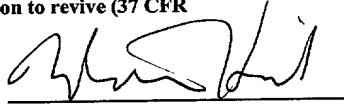


530 Rec'd PCT/PTC 28 JUN 2001

|   |   |   |   |
|---|---|---|---|
| FORM PTO-1390<br>(REV. 11-2000)   |   | U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE | ATTORNEY'S DOCKET NUMBER<br>19452A-002210US                         |
| TRANSMITTAL LETTER TO THE UNITED STATES<br>DESIGNATED/ELECTED OFFICE (DO/EO/US)<br>CONCERNING A FILING UNDER 35 U.S.C. 371  |   |   | U.S. APPLICATION NO. (if known, see 37 CFR 1.5)<br><b>09/869582</b> |
| INTERNATIONAL APPLICATION NO.<br>PCT/US99/24407   | INTERNATIONAL FILING DATE<br>October 15, 1999 | PRIORITY DATE CLAIMED<br>October 16, 1998               |   |
| TITLE OF INVENTION<br>METHODS OF SUPPRESSING FLOWERING IN TRANSGENIC PLANTS   |   |   |   |
| APPLICANT(S) FOR DO/EO/US<br>MARTIN F. YANOFISKY  |   |   |   |
| Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:   |   |   |   |
| <ol style="list-style-type: none"> <li>1. <input checked="" type="checkbox"/> This is a <b>FIRST</b> submission of items concerning a filing under 35 U.S.C. 371.</li> <li>2. <input type="checkbox"/> This is a <b>SECOND</b> or <b>SUBSEQUENT</b> submission of items concerning a filing under 36 U.S.C. 371.</li> <li>3. <input checked="" type="checkbox"/> This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (21) indicated below.</li> <li>4. <input checked="" type="checkbox"/> The US has been elected by the expiration of 19 months from the priority date (Article 31).</li> <li>5. <input checked="" type="checkbox"/> A copy of the International Application as filed (35 U.S.C. 37(c)(2)) <ol style="list-style-type: none"> <li>a. <input type="checkbox"/> is attached hereto (required only if not communicated by the International Bureau).</li> <li>b. <input checked="" type="checkbox"/> has been communicated by the International Bureau</li> <li>c. <input type="checkbox"/> is not required, as the application was filed in the United States Receiving Office (RO/US).</li> </ol> </li> <li>6. <input type="checkbox"/> An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)). <ol style="list-style-type: none"> <li>a. <input type="checkbox"/> is attached hereto.</li> <li>b. <input type="checkbox"/> has been previously submitted under 35 U.S.C. 154(d)(4).</li> </ol> </li> <li>7. <input checked="" type="checkbox"/> Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3)). <ol style="list-style-type: none"> <li>a. <input type="checkbox"/> are attached hereto (required only if not communicated by the International Bureau).</li> <li>b. <input type="checkbox"/> have been communicated by the International Bureau.</li> <li>c. <input type="checkbox"/> have not been made; however, the time limit for making such amendments has NOT expired.</li> <li>d. <input checked="" type="checkbox"/> have not been made and will not be made.</li> </ol> </li> <li>8. <input type="checkbox"/> An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).</li> <li>9. <input type="checkbox"/> An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).</li> <li>10. <input type="checkbox"/> An English language translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).</li> </ol> |   |   |   |
| Items 11 to 20 below concern document(s) or information included:   |   |   |   |
| <ol style="list-style-type: none"> <li>11. <input type="checkbox"/> An Information Disclosure Statement under 37 CFR 1.97 and 1.98.</li> <li>12. <input type="checkbox"/> An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.</li> <li>13. <input checked="" type="checkbox"/> A <b>FIRST</b> preliminary amendment.</li> <li>14. <input type="checkbox"/> A <b>SECOND</b> or <b>SUBSEQUENT</b> preliminary amendment.</li> <li>15. <input type="checkbox"/> A substitute specification.</li> <li>16. <input type="checkbox"/> A change of power of attorney and/or address letter.</li> <li>17. <input type="checkbox"/> A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825.</li> <li>18. <input type="checkbox"/> A second copy of the published international application under 36 U.S.C.</li> <li>19. <input type="checkbox"/> A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).</li> <li>20. <input checked="" type="checkbox"/> Other items or information:<br/>Copy of face sheet of published PCT application</li> </ol>  |   |   |   |

|  |  |  |
|--|--|--|
| I/S/ Application no. (if known, see 37 CFR 1.51)<br><div style="font-size: 2em; font-weight: bold;">09/869582</div>  | INTERNATIONAL APPLICATION NO<br>PCT/US99/24407 | ATTORNEY'S DOCKET NUMBER<br>19452A-002210US  |
| 21. <input checked="" type="checkbox"/> The following fees are submitted:  |  | CALCULATIONS PTO USE ONLY  |
| <b>BASIC NATIONAL FEE (37 CFR 1.492(A) (1) - (5)):</b><br>Neither international preliminary examination fee (37 CFR 1.492) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO .....\$1000.00<br><br>International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search report prepared by the EPO of JPO .....\$860.00<br><br>International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO .....\$710.00<br><br>International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) .....\$690.00<br><br>International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)(4) .....\$100.00 |  |  |
| <b>ENTER APPROPRIATE BASIC FEE AMOUNT =</b>  |  | \$860  |
| Surcharge of \$130.00 for furnishing the oath or declaration later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(e)).  |  | \$   |
| CLAIMS   | NUMBER FILED                                   | NUMBER EXTRA   |
| Total claims   | 33 - 20 =                                      | +13  |
| Independent claims   | 4 - 3 =  | +1   |
| MULTIPLE DEPENDENT CLAIM(S) (if applicable)  |  | + 270.00   |
| <b>TOTAL OF ABOVE CALCULATIONS =</b>   |  | \$1,174  |
| <input type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.   |  | \$   |
| <b>SUBTOTAL =</b>  |  | \$1,174  |
| Processing fee of \$130.00 for furnishing the English translation later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFT 1.492(f)).   |  | \$   |
| <b>TOTAL NATIONAL FEE =</b>  |  |  |
| Fee for recording the enclosed assignment (37 CFR 1.2(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). \$40.00 per property +  |  | \$   |
| <b>TOTAL FEES ENCLOSED =</b>   |  | \$1,174  |
|  |  | Amount to be refunded:   |
|  |  | \$   |
|  |  | charged:   |
|  |  | \$   |
| a. <input type="checkbox"/> A check in the amount of \$_____ to cover the above fees is enclosed.<br>b. <input checked="" type="checkbox"/> Please charge my Deposit Account No. <u>20-1430</u> in the amount of <u>\$1,174</u> to cover the above fees.<br>c. <input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. <u>20-1430</u> . A duplicate copy of this sheet is enclosed.<br>d. <input type="checkbox"/> Fees are to be charged to a credit card. <b>WARNING:</b> Information on this form may become public. <b>Credit card information should not be included on this form.</b> Provide credit card information and authorization on PTO-2038.   |  |  |
| NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b) must be filed and granted to restore the application to pending status.   |  |  |
| SEND ALL CORRESPONDENCE TO:  |  |  |
| Matthew E. Hinsch<br>Townsend and Townsend and Crew LLP<br>Two Embarcadero Center, 8th fl.<br>San Francisco, CA 94111  |  |  |
|  |  | <br>SIGNATURE |
|  |  | <u>Matthew E. Hinsch</u><br>NAME   |
|  |  | <u>47.651</u><br>REGISTRATION NUMBER   |

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re U.S. National Phase of  
PCT/US99/24407 of:

MARTIN F. YANOFSKY

Application No.: Not yet assigned

Filed: Herewith

For: METHODS OF SUPPRESSING  
FLOWERING IN TRANSGENIC  
PLANTS

PRELIMINARY AMENDMENT

San Francisco, CA 94111  
June 28, 2001

Box PCT  
Assistant Commissioner for Patents  
Washington, D.C. 20231

Sir:

Prior to the examination of the above-referenced application, please enter the following amendments and remarks.

IN THE CLAIMS:

Please substitute the following amended, clean version of the indicated claim (a marked-up version of the changes to the claim is attached to this Amendment):

8. (amended) A tissue derived from the transgenic plant of claim 1.

REMARKS:

Claims 1-33 are pending.

Amendment is made to delete the multiple dependency from claim 8, thereby  
avoiding the need to pay the multiple dependent surcharge.

Respectfully submitted,



Matthew E. Hinsch  
Reg. No. 47,651

TOWNSEND and TOWNSEND and CREW LLP  
Two Embarcadero Center, 8<sup>th</sup> Floor  
San Francisco, California 94111-3834  
Tel: (415) 576-0200  
Fax: (415) 576-0300  
MEH:tp

SF 1241620 v1

09869582.022802

**MARKED-UP VERSION OF THE CHANGES TO THE CLAIMS**

8. (amended) A tissue derived from the transgenic plant of [any of  
claims 1 to 7] claim 1.

098695137 020802

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

**PETITION FOR REVIVAL OF AN INTERNATIONAL APPLICATION FOR PATENT  
DESIGNATING THE U.S. ABANDONED UNINTENTIONALLY UNDER 37 CFR 1.137(b)**

Docket Number (Optional)

19452A-002210US

First named inventor: MARTIN F. YANOFSKY

Application No.: PCT/US99/24407

Group Art Unit:

Filed: October 15, 1999

Examiner:

Title: METHODS OF SUPPRESSING FLOWERING IN TRANSGENIC PLANTS

Attention: International Division, Legal Staff  
Box PCT  
Assistant Commissioner for Patents  
Washington, D.C. 20231**RECEIVED**  
17 AUG 2001  
Legal Staff  
International Division

The above-identified application became abandoned in the United States because the elements noted at 35 U.S.C. 371(c) were not filed prior to the expiration of the applicable time limit noted at 37 CFR 1.494(b) or (c) or 37 CFR 1.495(b) or (c). The date of abandonment is 04/17/01 (i.e., the day after the date on which the 35 U.S.C. 371(c) requirements were due; see 37 CFR 1.494(h) or 1.495(i)).

**APPLICANT HEREBY PETITIONS FOR REVIVAL OF THIS APPLICATION**

NOTE: A grantable petition requires the following items:

- (1) Petition fee
- (2) Proper response
- (3) Terminal disclaimer with disclaimer fee -- required for all applications filed before June 8, 1995; and
- (4) Statement that the entire delay was unintentional.

**1. Petition fee**

- ☐ Small entity - fee \$ \_\_\_\_\_ (37 CFR 1.17(m))
- ☐ Small entity statement enclosed herewith.
- ☐ Small entity statement previously filed.
- ☒ Other than small entity - fee \$ 1,240 (37 CFR 1.17(m))

**2. Proper response**

A. The proper response (the missing 35 U.S.C. 371(c) requirements) in the form of  
U.S. National Phase filing (identify type of response):

- ☐ has been filed previously on \_\_\_\_\_
- ☒ is enclosed herewith.

7/05/2001 ATRAM1 00000136 201430 09869582

4 FC:141

1240.00 CH

[Page 1 of 2]

Burden Hour Statement: This form is estimated to take 1.0 hour to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

2002220-23559850

3. Terminal disclaimer with disclaimer fee

☒ Since this utility/plant application was filed on or after June 8, 1995, no terminal disclaimer is required.

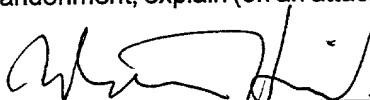
☐ A terminal disclaimer (and disclaimer fee (37 CFR 1.20(d)) of \$\_\_\_\_\_ for a small entity or \$\_\_\_\_\_ for other than a small entity) equivalent to the number of months from abandonment to the filing of this petition is enclosed herewith.

4. Statement. The entire delay in filing the 35 U.S.C. 371(c) requirements from their due date until the filing of a grantable petition under 37 CFR 1.137(b) was unintentional.

Where a petition under 37 CFR 1.137(b) is not filed within three months from the mail date of any notice of abandonment or one year from the date of abandonment, explain (on an attached sheet) in detail the cause of the delay in filing this petition.

June 28, 2001

Date



Signature

Telephone

Number: (415) 576-0200

Matthew E. Hinsch 47,651

Typed or printed name

Townsend and Townsend and Crew LLP

Address

Two Embarcadero Center, 8th Fl.

San Francisco, CA 94111

Enclosures: ☒ Response

☒ Fee Payment

☐ Terminal Disclaimer Form

☐ Small Entity Status Form

☐ \_\_\_\_\_

## **Application Data Sheet**

### **Application Information**

Application number:: 09/869,582

Filing Date::

Application Type:: Regular

Subject Matter:: Utility

Suggested classification::

Suggested Group Art Unit::

Sequence Submission::

Computer Readable Form (CRF)?::

Number of copies of CRF::

Title:: METHODS OF SUPPRESSING FLOWERING IN  
TRANSGENIC PLANTS

Attorney Docket Number:: 19452A-002210US

Request for Early Publication:: No

Request for Non-Publication:: No

Suggested Drawing Figure::

Total Drawing Sheets:: 43

Small Entity?:: No

Latin name::

Variety denomination name::

Petition included?:: No

Petition Type::

Licensed US Govt. Agency::

Contract or Grant Numbers One::

Secrecy Order in Parent Appl.: No

### **Applicant Information**

Applicant Authority Type:: Inventor

Primary Citizenship Country:: US



Status:: Full Capacity  
Given Name:: Martin  
Middle Name:: E.  
Family Name:: Yanofsky  
Name Suffix::  
City of Residence:: San Diego CA  
State or Province of Residence:: CA  
Country of Residence:: US  
Street of Mailing Address:: 5039 Manor Ridge Lane  
City of Mailing Address:: San Diego  
State or Province of mailing address:: CA  
Country of mailing address::  
Postal or Zip Code of mailing address:: 92130

**Correspondence Information**

Correspondence Customer Number:: 20350

**Representative Information**

Representative Customer Number:: 20350

**Domestic Priority Information**

|                       |                    |                      |                      |
|-----------------------|--------------------|----------------------|----------------------|
| Application::         | Continuity Type::  | Parent Application:: | Parent Filing Date:: |
| This application is a | 371 of             | PCT/US99/24407       | 10/15/99             |
| which                 | claims priority of | 60/104,604           | 10/16/98             |

**Foreign Priority Information**

|           |                      |               |
|-----------|----------------------|---------------|
| Country:: | Application number:: | Filing Date:: |
|-----------|----------------------|---------------|

### **Assignee Information**

Assignee Name:: The Regents of the University of California  
Street of mailing address:: 1111 Franklin Street, 12th Floor  
City of mailing address:: Oakland  
State or Province of mailing address:: CA  
Country of mailing address:: USA  
Postal or Zip Code of mailing address:: 94607

METHODS OF SUPPRESSING FLOWERING IN TRANSGENIC PLANTSFIELD OF THE INVENTION

5       The present invention relates generally to plant molecular biology and genetic engineering and more specifically to the production of genetically modified plants in which the natural process of flowering is suppressed.

BACKGROUND INFORMATION

10       The ecological and economic importance of wood is difficult to overstate, with the total amount of wood in the world's forests estimated at about 1.5 Gt. Thus, wood is by far the most abundant component of the terrestrial biomass. The carbon stored in wood and humus (partially degraded wood) is important in the planetary carbon cycle, which has a significant influence on global climate. In addition, wood is a leading industrial component of the global economy. About 4% of the US gross national product has been attributed to the wood  
15       products industry in past decades.

      Unfortunately, a growing population is reducing the arable land area in the United States and around the world, while the demand for wood products increases. This growing demand and limited resources have resulted in a need for greater productivity of the remaining forest lands.

20       The flowering process consumes 25 to 35% of the energy of a typical plant, thereby limiting wood production. Thus, for trees used for lumber or pulp production, for example, it can be advantageous to suppress flowering in order increase the yield of wood. Suppression of flowering also can be desired to eliminate the production of allergic pollen, or to prevent pollen dissemination. Unfortunately, methods of producing genetically modified plants in  
25       which flowering is suppressed without effecting other desirable traits are not currently available.

      Thus, a need exists for developing genetically modified plant varieties in which the natural process of flowering is suppressed. The present invention satisfies this need and provides related advantages as well.

SUMMARY OF THE INVENTION

The present invention provides a transgenic plant characterized by suppressed flowering. The transgenic plant contains a nucleic acid molecule including a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product, wherein the nucleic acid molecule is heritable by progeny thereof.

The transgenic plant contains a nucleic acid molecule including a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product, where the floral organ selective regulatory element is an *AGL2* regulatory element, an *AGL4* regulatory element or an *AGL9* regulatory element, or a *API* regulatory element, and wherein the nucleic acid molecule is heritable by progeny thereof.

In a transgenic plant of the invention, the floral organ selective regulatory element can be, for example, an *AGL2* regulatory element having substantially the nucleotide sequence of *Arabidopsis AGL2* promoter SEQ ID NO:1, or an active fragment thereof. A floral organ selective regulatory element useful in a transgenic plant of the invention also can be, for example, an *AGL4* regulatory element such as an *AGL4* regulatory element having substantially the nucleotide sequence of *Arabidopsis AGL4* promoter SEQ ID NO:2, or an active fragment thereof. A floral organ selective regulatory element also can be an *AGL9* regulatory element such as an *AGL9* regulatory element having substantially the nucleotide sequence of *Arabidopsis AGL9* promoter SEQ ID NO:3, or an active fragment thereof. A floral organ selective regulatory element also can be an *API* regulatory element such as an *API* regulatory element having substantially the nucleotide sequence of *Arabidopsis API* promoter SEQ ID NO:10, or an active fragment thereof.

DNA sequences encoding a variety of encoded cytotoxic gene products can be used to produce a transgenic plant of the invention, including DNA encoding toxic peptides such as the diphtheria toxin A chain, RNase T1, Barnase RNase, ricin toxin A chain or the herpes simplex virus thymidine kinase (tk) gene product.

The invention further relates to regenerated fertile seedlings and mature plants obtained from transgenic seed or from the vegetative reproduction of transgenic plants, and R1 and subsequent generations, produced by sexual propagation or vegetative reproduction.

The description of the invention hereafter refers to *Arabidopsis thaliana*, when necessary for the sake of example. However, it should be noted that the invention is not limited to genetic transformation of plants such as *Arabidopsis*. The method of the present invention is capable of being practiced for other plant species, including for example, other

The present invention also provides a tissue derived from a transgenic plant characterized by suppressed flowering and containing a nucleic acid molecule including a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product, wherein the nucleic acid molecule is heritable by progeny thereof.

The invention additionally is directed to all products derived from transgenic plants, plant cells, plant parts and seeds, which contain a nucleic acid molecule including a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product, wherein the nucleic acid molecule is heritable by progeny thereof.

Also provided by the present invention is a method of producing a fertile, transgenic plant characterized by suppressed flowering. The method is based upon transformation of plant material, selection, plant regeneration, and conventional or propagation breeding techniques.

30 The method includes the step of introducing into a plant an exogenous nucleic acid molecule containing a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product (a peptide), wherein the nucleic acid molecule is heritable by asexual or sexually obtained progeny thereof. The method includes

the step of introducing into a plant an exogenous nucleic acid molecule containing a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product, where flowering is suppressed due to selective expression of the exogenous nucleic acid molecule and where the floral organ selective regulatory element is preferably an *AGL2* regulatory element, an *AGL4* regulatory element or an *AGL9* regulatory element, or the *API* regulatory element.

The present invention also provides an isolated nucleic acid molecule including an *AGL2*, *AGL4* or *AGL9* or *API* regulatory element, which confers selective expression upon an operatively linked nucleotide sequence (structural gene) in one or more floral organs of a plant.

The isolated nucleic acid molecule can further include, if desired, an operatively linked nucleotide sequence encoding a cytotoxic gene product. The encoded cytotoxic gene product can be one of a variety of cytotoxic gene products such as the peptides diphtheria toxin A chain, RNase T1, Barnase RNase, ricin toxin A chain or herpes simplex virus thymidine kinase gene product.

The present invention also provides a kit for producing a transgenic plant characterized by suppressed flowering. A kit of the invention comprises packaging containing a plant expression vector comprising a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product, and instructions for transforming a susceptible plant with said vector.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1a through 1e shows the *Arabidopsis AGL2* promoter SEQ ID NO:1.

Figure 2a through 2f shows the *Arabidopsis AGL4* promoter SEQ ID NO:2.

Figure 3a through 3q shows the *Arabidopsis AGL9* promoter SEQ ID NO:3.

Figure 4 shows the nucleotide (SEQ ID NO:4) and amino acid sequence (SEQ ID NO:5) of the *AGL2* cDNA and the nucleotide (SEQ ID NO:6) and amino acid sequence (SEQ ID NO:7) of the *AGL4* cDNA. The *AGL2* sequences are shown above the *AGL4* sequences.

Figure 5 shows the nucleotide (SEQ ID NO:8) and deduced amino acid sequence (SEQ ID NO:9) of the *AGL9* cDNA.

Figure 6a through 6f shows the *Arabidopsis API* promoter SEQ ID NO: 10.

Figure 7 shows a diagram of reporter construct POP10. The construct has 1.7 kb *API* promoter plus the entire coding region of *API* in front of promoterless GUS gene in pBI101.2

plasmid. The construct has 1.7 kb *API* promoter plus the entire coding region of *API* in front of promoterless GUS gene in pBI101.2 plasmid. The construct was first made by PCR amplification from intron 3 to the end of *API* gene in exon 8 (right before stop codon) using KY65 plasmid containing *API* genomic region as template. The HindIII site was added to the forward primer AP1HIN [5'-CAAGCTTGTACACATTTACTCATCACAT-3'] and BamHI site was added to reverse primer AP1BAM, [5'-CGGATCCTGCGCGAAGCAGCCAAGGTTG-3'] to aid cloning (sequence in *italics* are restriction sites of HindIII and BamHI). The 1.7 kb amplified fragment was cloned into plasmid pBI101.2 using HindIII and BamHI sites giving construct POP9. The 3.6 kb HindIII / XbaI fragment was isolated from KY65 plasmid and cloned into POP9 construct giving POP10 construct.

Figure 8a through 8b shows the nucleotide (SEQ ID NO:11) and deduced amino acid sequence (SEQ ID NO:12) of the *API* cDNA.

Figure 9 shows GUS expression in 2 representative *API* reporter lines. GUS activity is flower specific and GUS staining pattern largely mimics *API* RNA accumulation pattern.

Figure 10a through 10b shows the nucleotide (SEQ ID NO:6) and amino acid sequence (SEQ ID NO:7) of the *AGL4* cDNA.

Figure 11a through 11b shows the nucleotide (SEQ ID NO:4) and amino acid sequence (SEQ ID NO:5) of the *AGL2* cDNA.

#### DETAILED DESCRIPTION OF THE INVENTION

Flowering is often desirable and is the natural mechanism by which flowering plants propagate. Yet for some applications, it can be desirable to suppress flower and seed production. For example, in trees grown for lumber or pulp, wood yield can be increased by suppressing flower and seed production, which normally consumes 25 to 35% of the energy of a typical plant. Where allergic pollens are a concern, non-flowering varieties are desirable to avoid pollen dissemination. Furthermore, flowering can hasten senescence; thus, non-flowering transgenic plants can have improved longevity.

The present invention provides transgenic plants characterized by suppressed flowering. In a transgenic plant of the invention, a regulatory element that directs selective expression in one or more floral organs is used to control expression of an inhibitory or cytotoxic peptide such as diphtheria toxin or ricin. The selectively expressed cytotoxic gene product destroys

floral tissue, thereby suppressing flowering, but is not expressed significantly in vegetative or other tissues and so has no deleterious effect outside the floral tissue.

A fertile transgenic plant of the invention contains a nucleic acid molecule including a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product, wherein the nucleic acid molecule is heritable by progeny thereof.

5 A fertile transgenic plant of the invention contains a nucleic acid molecule including a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product, the floral organ selective regulatory element is an *AGL2* regulatory element, an *AGL4* regulatory element or an *AGL9* regulatory element or an *API* regulatory element, wherein the nucleic acid molecule is heritable by progeny thereof.

10

"Transgenic" is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered beneficially by the presence of heterologous DNA that was introduced into the genotype by a process of genetic engineering, or which was initially introduced into the genotype of a parent plant by such a process and is subsequently transferred to later generations by sexual or asexual cell crosses or cell divisions. As used

15 herein, "genotype" refers to the sum total of genetic material within a cell, either chromosomally, or extrachromosomally borne. Therefore, the term "transgenic" as used herein does not encompass the alteration of the genotype of any plant by conventional plant breeding methods or by naturally occurring events such as random cross-fertilization or spontaneous mutation.

20

The term "transgenic" may be used herein to describe a plant that contains an exogenous nucleic acid molecule or chimeric nucleic acid construct, which can be derived from an orthologous or heterologous plant or can originate from an animal or virus.

The term "exogenous," as used herein in reference to a nucleic acid molecule and a transgenic plant, means a nucleic acid molecule that is not native to the plant or that is present in the genome in other than its native association. An exogenous nucleic acid molecule can have a naturally occurring or non-naturally occurring nucleotide sequence and can be orthologous or heterologous to the plant species into which it is introduced.

25

The term "heritable" refers to the fact that the nucleic acid molecule is capable of transmission through a complete sexual cycle of a plant, i.e., it is passed from one plant through its gametes to progeny plants in the same manner as occurs in normal plants, or the nucleic acid can be transmitted via asexual propagation of cuttings or shoots.

30



The term "operatively linked," as used in reference to a regulatory element and a nucleotide sequence encoding a cytotoxic gene product, means that the regulatory element is linked so that it confers regulated expression upon the operatively linked nucleotide sequence. Thus, the term "operatively linked," as used in reference to a floral organ selective regulatory element and a nucleotide sequence encoding a cytotoxic gene product, means that the floral organ selective regulatory element is linked to the nucleotide sequence encoding the cytotoxic gene product so that the expression pattern of the floral organ selective regulatory element is conferred upon the nucleotide sequence encoding the cytotoxic gene product. It is recognized that a regulatory element and a nucleotide sequence that are operatively linked have, at a minimum, all elements essential for transcription, including, for example, a TATA box.

The term "suppressed," as used herein in reference to the flowering of a transgenic plant of the invention, means a significantly diminished extent of flowering as compared to the extent of flowering in a corresponding plant lacking a nucleic acid molecule containing a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product. Thus, the term "suppressed" is used broadly to encompass both flowering that is significantly reduced as compared to the flowering in a corresponding non-transgenic plant, and to flowering that is completely precluded. In view of the above, one skilled in the art recognizes that a transgenic plant of the invention can be completely sterile or can be characterized by reduced fertility although generally flowering is suppressed to the extent that the transgenic plant is completely sterile.

Two amino acid sequences are homologous if there is a partial or complete identity between their sequences. For example, 85% homology means that 85% of the amino acids are identical when the two sequences are aligned for maximum matching. Gaps (in either of the two sequences being matched) are allowed in maximizing matching; gap lengths of 5 or less are preferred with 2 or less being more preferred. Alternatively and preferably, two protein sequences (or polypeptide sequences derived from them of at least 30 amino acids in length) are homologous, as this term is used herein, if they have an alignment score of at more than 5 (in standard deviation units) using the program ALIGN with the mutation data matrix and a gap penalty of 6 or greater. See Dayhoff, M. O., in Atlas of Protein Sequence and Structure, 1972, volume 5, National Biomedical Research Foundation, pp. 101-110, and Supplement 2 to this volume, pp. 1-10. The two sequences or parts thereof are more

preferably homologous if their amino acids are greater than or equal to 50% identical when optimally aligned using the ALIGN program.

As used herein, the term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The terms "substantial identity" as used herein denote a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 85 percent sequence identity, preferably at least 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 20-50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. The reference sequence may be a subset of a larger sequence, for example, as a segment of human MCP-1.

As used herein, the term "flowering" is used broadly to refer not only to the traditional flowering of angiosperms but also to the normal reproductive development of other plants such as conifers.

It is recognized that there can be natural variation in the extent of flowering within a plant species or variety. However, a "suppression" in flowering in a transgenic plant of the invention readily can be identified by sampling a population of the corresponding plants, such as wild type plants, and determining that the normal distribution of flowering is significant diminished, on average, as compared to the normal distribution of flowering in a population of the corresponding plant species or variety that does not have a nucleic acid molecule containing a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product. Thus, production of transgenic plants of the invention provides a means to skew the extent of normal flowering, such that flowering is

diminished, on average, at least about 1%, 2%, 5%, 10%, 30%, 50% or 100% as compared to flowering in the corresponding plant species that does not have a nucleic acid molecule containing a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product.

5 As used herein, the term "cytotoxic gene product" means a gene product, usually a peptide, that inhibits the growth of, or causes the death of, the cell in which it is expressed. Preferably, a cytotoxic gene product does not result in the death of cells other than the cell in which it is expressed. Thus, expression of a cytotoxic gene product from a floral organ selective regulatory element can be used to ablate cells within one or more floral organs  
10 without disturbing neighboring cells. A variety of cytotoxic gene products useful in plants are known in the art including toxins and enzymes, for example, diphtheria toxin A chain polypeptides; RNase T1; Barnase RNase; ricin toxin A chain polypeptides; and herpes simplex virus thymidine kinase (tk) gene products. While the diphtheria toxin A chain, RNase T1 and Barnase RNase are preferred cytotoxic gene products, or multiple nucleotide  
15 sequences encoding other cytotoxic gene products, can be used with a floral organ selective regulatory element to generate a transgenic plant of the invention characterized by suppressed flowering.

Diphtheria toxin is the naturally occurring toxin of *Cornebacterium diphtheriae*, which catalyzes the ADP-ribosylation of elongation factor 2, resulting in inhibition of protein  
20 synthesis and consequent cell death (Collier, Bacteriol. Rev. 39:54-85 (1975)). A single molecule of the fully active toxin is sufficient to kill a cell (Yamaizumi et al., Cell 15:245-250 (1978)). Diphtheria toxin has two subunits: the diphtheria toxin B chain directs internalization to most eukaryotic cells through a specific membrane receptor, whereas the A chain encodes the toxic catalytic domain. The catalytic DT-A chain does not include a signal  
25 peptide and is not secreted. Further, any DT-A released from dead cells in the absence of the diphtheria toxin B chain is precluded from cell attachment. Thus, DT-A is cell autonomous and directs killing only of the cells in which it is expressed without apparent damage to neighboring cells. The DT-A expression cassette of Palmiter et al., which contains the 193 residues of the A chain engineered with a synthetic ATG and lacking the native leader  
30 sequence, is particularly useful in the transgenic plants of the invention (Palmiter et al., Cell 50:435-443 (1987); Greenfield et al., Proc. Natl. Acad. Sci., USA 80:6853-6857 (1983), each of which is incorporated herein by reference).

RNase T1 of *Aspergillus oryzae* and Barnase RNase of *Bacillus amylolique-faciens* also are cytotoxic gene products useful in the transgenic plants of the invention (Thorsness and Nasrallah, Methods in Cell Biology 50:439-448 (1995)). Barnase RNase may be more generally toxic to plants than RNase T1 and, thus, is preferred in the methods of the invention.

Ricin, a ribosome-inactivating protein produced by castor bean seeds, also is a cytotoxic gene product useful in a transgenic plant of the invention. The ricin toxin A chain polypeptide can be used to direct cell-specific ablation as described, for example, in Moffat et al., Development 114:681-687 (1992). Plant ribosomes are variably susceptible to the plant-derived ricin toxin. The skilled person understands that the toxicity of ricin depends is variable and should be assessed for toxicity in the plant species of interest (see Olsnes and Pihl, Molecular Action of Toxins and Viruses, pages 51-105, Amsterdam: Elsevier Biomedical Press (1982)).

The present invention relates to the use of floral organ selective regulatory elements derived from *AGL2*, *AGL4* or *AGL9*, which are "*AGAMOUS-LIKE*" or "*AGL*" genes. *AGAMOUS* (*AG*) is a floral organ identity gene, one of a related family of transcription factors that, in various combinations, specify the identity of the floral organs: the petals, sepals, stamens and carpels (Bowman et al., Devel. 112:1-20 (1991); Weigel and Meyerowitz, Cell 78:203-209 (1994); Yanofsky, Annual Rev. Plant Physiol. Mol. Biol. 46:167-188 (1995)). The *AGAMOUS* gene product is essential for specification of carpel and stamen identity (Bowman et al., The Plant Cell 1:37-52 (1989); Yanofsky et al., Nature 346:35-39 (1990)). Related genes have recently been identified and denoted "*AGAMOUS-LIKE*" or "*AGL*" genes (Ma et al., Genes Devel. 5:484-495 (1991); Mandel and Yanofsky, The Plant Cell 7:1763-1771 (1995), which is incorporated herein by reference).

*AGL2*, *AGL4* and *AGL9*, like *AGAMOUS* and other *AGL* genes, are characterized, in part, in that each is a plant MADS box gene. The plant MADS box genes generally encode proteins of about 260 amino acids including a highly conserved MADS domain of about 56 amino acids (Riechmann and Meyerowitz, Biol. Chem. 378:1079-1101 (1997), which is incorporated herein by reference). The MADS domain, which was first identified in the *Arabidopsis AGAMOUS* and *Antirrhinum majus DEFICIENS* genes, is conserved among transcription factors found in humans (serum response factor; SRF) and yeast (MCM1; Norman et al., Cell 55:989-1003 (1988); Passmore et al., J. Mol. Biol. 204:593-606 (1988), and is the most highly conserved region of the MADS domain proteins. The MADS domain

is the major determinant of sequence specific DNA-binding activity and can also perform dimerization and other accessory functions (Huang et al., The Plant Cell 8:81-94 (1996)). The MADS domain frequently resides at the amino-terminus, although some proteins contain additional residues amino-terminal to the MADS domain.

5 The "intervening domain" or "I-domain," located immediately C-terminal to the MADS domain, is a weakly conserved domain having a variable length of approximately 30 amino acids (Purugganan et al., Genetics 140:345-356 (1995)). In some proteins, the I-domain plays a role in the formation of DNA-binding dimers. A third domain present in plant MADS domain proteins is a moderately conserved 70 amino acid region denoted the "keratin-like domain" or "K-domain." Named for its similarity to regions of the keratin molecule, the structure of the K-domain appears capable of forming amphipathic helices and may mediate protein-protein interactions (Ma et al., Genes Devel. 5:484-495 (1991)). The most variable domain, both in sequence and in length, is the carboxy-terminal or "C-domain" of the MADS domain proteins. Dispensable for DNA binding and protein dimerization in some MADS domain proteins, the function of the C-domain remains unknown.

10 The amino acid sequence of *Arabidopsis AGL2*, a protein with a calculated molecular mass of about 28.5 kDa, is shown in Figures 4 and 11a through 11b. Like other AGAMOUS-LIKE proteins, *AGL2* has a highly conserved MADS domain and a K domain (Ma et al., Genes Devel. 5:484-495 (1991)). RNA dot blot hybridization was used to analyze *AGL2* expression in immature seed pods, flowers, stems, and leaves. *AGL2* RNA was preferentially expressed in flowers: a strong hybridization signal was seen in flower RNA, with a diminished level seen in RNA from immature seed pods. A faint signal was also detected in leaves. To determine whether *AGL2* is expressed in an organ-specific manner, *in situ* hybridization was performed with wild type *Arabidopsis* inflorescence sections. The results showed that *AGL2* was expressed mainly in carpels and was concentrated there in the ovules. In addition, *AGL2* was expressed at a lower level in the stamens, with expression restricted to the anthers. Thus, the *AGL2* gene is selectively expressed in floral organs, with a high level of expression seen in flowers and young seed pods and a much lower level of expression seen in leaves. These results indicate that an *AGL2* regulatory element can confer floral organ selective expression upon a heterologous linked gene.

25 The amino acid sequence of *AGL4* is shown in Figures 4 and 10a through 10b. The encoded protein, which has a calculated molecular mass of 28.5 kDa, has the characteristic highly conserved MADS domain. RNA dot blot hybridization was used to assess *AGL4*

expression in immature seed pods, flowers, stems, and leaves. *AGL4* was highly expressed in flowers with the expression continuing at a lower level in immature seed pods. No expression was seen in the vegetative stems and leaves. These results indicate that *AGL4* is specifically expressed in flowers and that an *AGL4* regulatory element can confer floral organ selective expression upon a heterologous linked gene.

*Arabidopsis AGL9* is a 251 amino acid protein having a calculated molecular mass of 29 kDa. *AGL9* has a highly conserved MADS domain, as well as a K domain (see Figure 5). The protein encoded by *Arabidopsis AGL9* has a high degree of similarity to the products of the *TM5* gene from tomato (*Lycopersicon esculentum*); the petunia gene *FBP2*, and the *DEFH200* gene from *Antirrhinum majus*, indicating that *TM5*, *FBP2* and *DEFH200* are *AGL9* orthologs (Pnueli et al., *Plant J.* 1:255-266 (1991); Angenent et al., *Plant Cell* 4:983-993 (1992); and Davies et al., *EMBO J.* 15:4330-4343 (1996), each of which is incorporated herein by reference). Throughout the first 160 amino acids, *AGL9* shares approximately 89% amino acid identity with the *FBP2*, *TM5* and *DEFH200* gene products.

*AGL9* RNA accumulates only in flowers, with RNA blot analysis showing no detectable expression in roots, stems or cauline leaves. *In situ* hybridization analyses demonstrated that *AGL9* RNA begins to accumulate after the onset of expression of the floral meristem identity genes but before the expression of the floral organ identity genes. In particular, floral meristem identity genes such as *API* and *CAL* are first expressed during stage 1 flower primordia, followed by *AGL2* and *AGL4*, which are first expressed throughout stage 2 flower primordia. *AGL9* is subsequently expressed late in stage 2 in a region that does not include the outer perimeter of the flower primordium. Later in flower development, *AGL9* RNA accumulates in the petal, stamen, and carpel organs. Thus, *AGL9* is specifically expressed only in floral organs, indicating that an *AGL9* regulatory element can confer floral organ selective expression upon a heterologous linked gene.

The amino acid sequence of *API* is shown in Figure 8a through 8b (Mandel, 1992 *Nature* 360:273-277). The encoded protein, which has a calculated molecular mass of 30 kDa, has the characteristic highly conserved MADS domain. The deduced *API* protein is similar to the snapdragon *SQUAMOSA* protein, sharing 68% identical amino acid residues (Huijser et al., *EMBO J.* 33:1239-1249; 1992). RNA blot hybridization was used to assess *API* expression in roots, stems, leaves, and flowers, where it was shown to be flower specific (Id., Figure 3). Subsequent RNA tissue in situ hybridizations further defined the *API* RNA accumulation pattern where it was shown to first

be expressed in a young flower primordium (a flower meristem) when it first becomes visible on the flanks of the shoot meristem. Additional studies showed that *API* RNA accumulates in all cells of the young flower, and that in mature flowers, *API* is expressed in sepals and petals but not in stamens and carpels (Id., Fig. 4). Thus, *API* is specifically expressed in flowers and that an *API* regulatory element can confer floral organ selective expression upon a heterologous linked gene. Proof of this concept came from fusing the *API* regulatory region to the easily assayable "GUS" marker gene and the subsequent generation of transgenic plants that had stably integrated the *API*::GUS transgene into the plant nuclear genome (the POP10 construct and resulting lines)(See Figure 9).

The *API* regulatory region includes the 1.7 kb of the *API* "promoter" (the promoter is defined as the 1700 bp immediately upstream of the *API* translation initiation codon, ATG), as well as the genomic region containing all *API* intronic sequences. Both the "full length" *API* promoter (*API* promoter plus all genomic regions containing *API* intronic sequences as shown for the POP10 construct in Figure 7) and the 1700 bp *API* promoter fragment are sufficient to express foreign genes that are operably linked to it within flowers, and thus may be suitable for suppressing flowering. Smaller constructs, such as those that do not contain all of the *API* intronic sequences, may also be flower specific, and thus it is not necessary to include all of the *API* genomic sequences to achieve complete flower-specific regulation. However, the use of the "full length" *API* regulatory region may be used for optimal flower specific expression, since these sequences will drive gene expression only in flowers.

As used herein, the term "floral organ selective regulatory element" refers to a regulatory element such as a 5', 3' or intronic regulatory element that, when operatively linked to a nucleotide sequence, confers selective expression upon the operatively linked nucleotide sequence in a limited number of plant tissues, including one or more floral organs or subparts thereof. Thus, a floral organ selective regulatory element, as defined herein, confers selective expression in the petals, sepals, stamens or carpels of a plant or in some cell types within the petals, sepals, stamens or carpels, with expression low or absent in other tissues of the plant.

A floral organ selective regulatory element can confer specific expression exclusively in cells of one or more floral organ, or can confer selective expression in a limited number of plant cell types including cells of one or more floral organ. For example, an *AGL9* regulatory element, which confers specific expression in flowers, without conferring expression in vegetative tissues such as roots, stems or cauline leaves, is a floral organ selective regulatory

element as defined herein. A floral organ selective regulatory element also can be, for example, an *AGL2* regulatory element, which confers high level expression in flowers, with a minimal level of expression in leaves.

As used herein, the term "*AGL2* regulatory element" refers to a regulatory element  
5 derived from *Arabidopsis AGL2* (SEQ ID NO:5) or an ortholog of *Arabidopsis AGL2*. An  
*AGL2* ortholog is a MADS box gene product expressed, at least in part, in one or more floral  
organs of a plant and having homology to the amino acid sequence of *Arabidopsis AGL2*  
(SEQ ID NO:5). An *AGL2* ortholog can be, for example, a pine or rice ortholog such as  
PrMADS1 or OsMADS5 (Mouradov et al., *Plant Physiol.* 117:55-62 (1998); Kang and An,  
10 *Mol. Cells* 7:45-51 (1997), each of which is incorporated herein by reference) or can be  
another ortholog such as a *Eucalyptus* or spruce ortholog. An *AGL2* ortholog generally has at  
least about 80% amino acid identity with amino acids 1 to 160 of *Arabidopsis AGL2* (SEQ ID  
NO:5) and can have, for example, at least about 85%, 90%, or 95% amino acid identity with  
amino acids 1 to 160 of *Arabidopsis AGL2* (SEQ ID NO:5).

As used herein, the term "*AGL4* regulatory element" refers to a regulatory element  
15 derived from *Arabidopsis AGL4* (SEQ ID NO:7) or an ortholog of *Arabidopsis AGL4*. An  
*AGL4* ortholog is a MADS box gene product expressed, at least in part, in one or more floral  
organs of a plant and having homology to the amino acid sequence of *Arabidopsis AGL4*  
(SEQ ID NO:7). An *AGL4* ortholog can be, for example, a *Eucalyptus*, pine or spruce  
20 ortholog. An *AGL4* ortholog generally has at least about 80% amino acid identity with amino  
acids 1 to 160 of *Arabidopsis AGL4* (SEQ ID NO:7) and can have, for example, at least about  
85%, 90%, or 95% amino acid identity with amino acids 1 to 160 of *Arabidopsis AGL4* (SEQ  
ID NO:7).

As used herein, the term "*AGL9* regulatory element" refers to a regulatory element  
25 derived from *Arabidopsis AGL9* (SEQ ID NO:9) or an ortholog of *Arabidopsis AGL9*. An  
*AGL9* ortholog is a MADS box gene product expressed, at least in part, in one or more floral  
organs of a plant and having homology to the amino acid sequence of *Arabidopsis AGL9*  
(SEQ ID NO:9). An *AGL9* ortholog can be, for example, a tomato, petunia or *A. majus*  
ortholog such as TM5, FBP2 or DEFH200 (Pnueli et al., *The Plant Cell* 6:163-173 (1994);  
30 Angenent et al., *Plant Cell* 4:983-993 (1992); and Davies et al., *EMBO J.* 15:4330-4343  
(1996)) or can be, for example, a *Eucalyptus*, pine or spruce ortholog. An *AGL9* ortholog  
generally has at least about 80% amino acid identity with amino acids 1 to 160 of



*Arabidopsis AGL9* (SEQ ID NO:9) and can have, for example, at least about 85%, 90%, or 95% amino acid identity with amino acids 1 to 160 of *Arabidopsis AGL9* (SEQ ID NO:9).

As used herein the term "*AP1* regulatory element " refers to a regulatory element derived from *Arabidopsis AP1* (SEQ ID NO:10) or an ortholog of *Arabidopsis AP1*. An *AP1* ortholog is a MADS box gene product expressed, at least in part, in one or more floral organs of a plant and having homology to the amino acid sequence of *Arabidopsis AP1* (SEQ ID NO:10). An *AP1* ortholog can be, for example, a snapdragon ortholog, such as SQUAMOSA. Also, an *AP1* ortholog could be, for example, a *Eucalyptus*, pine or spruce ortholog. An *AP1* ortholog generally has at least about 75% amino acid identity with amino acids 1 to 160 of *Arabidopsis AP1* (SEQ ID NO:10) and can have, for example, at least about 85%, 90%, or 95% amino acid identity with amino acids 1 to 160 of *Arabidopsis AP1* (SEQ ID NO:10).

Preferably, an *AGL2*, *AGL4* or *AGL9* or *AP1* floral organ selective regulatory element is orthologous to the transgenic plant species into which it is introduced. An *AGL2* promoter (SEQ ID NO:1) or active fragment thereof, for example, can be introduced into an *Arabidopsis* plant to produce a transgenic *Arabidopsis* variety characterized by suppressed flowering. Similarly, a *Eucalyptus AGL2*, *AGL4* or *AGL9* or *AP1* floral organ selective regulatory element can be introduced into a *Eucalyptus* plant to produce a transgenic *Eucalyptus* variety characterized by suppressed flowering.

An *AGL2*, *AGL4* or *AGL9* or *AP1* floral organ selective regulatory element also can be introduced into a heterologous plant to produce a transgenic plant of the invention characterized by suppressed flowering. AGAMOUS-like gene products have been widely conserved throughout the plant kingdom; for example, AGAMOUS has been conserved in tomato (TAG1) and maize (ZAG1), indicating that orthologs of AGAMOUS-like genes are present in most, if not all, angiosperms (Pnueli et al., The Plant Cell 6:163-173 (1994); Schmidt et al., The Plant Cell 5:729-737 (1993)). Furthermore, it has been shown that MADS-box genes exist in gymnosperms and angiosperms as well as in ferns, the common ancestors of contemporary seed plants (Tandre et al., Plant Mol. Biol. 27:69-78 (1995); Liu and Podila, Plant Phys. 113:665 (1997); Münster et al., Proc. Natl. Acad. Sci., USA 94:2145-2420 (1997); and Mouradov et al., Plant Physiol. 117:55-62 (1998)). *AGL2*, *AGL4* and *AGL9* floral organ selective regulatory elements also can be conserved and can function across species boundaries to confer floral organ selective expression in heterologous plant species. Thus, an *Arabidopsis AGL2*, *AGL4* or *AGL9* or *AP1* floral organ selective regulatory

element, such as the *Arabidopsis* AGL2, AGL4 or AGL9 or API promoter SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3 or SEQ ID NO: 10, or an active fragment thereof, can confer floral organ selective expression upon an operatively linked nucleotide sequence encoding a cytotoxic gene product in a heterologous plant such as *Eucalyptus*, whereby the cytotoxic gene product is selectively expressed in floral tissue and flowering is suppressed.

A transgenic plant of the invention that is characterized by suppressed flowering can be one of a variety of plant species. As used herein, the term "plant" means a higher plant that generally is a vascular plant or seed plant such as an angiosperm or gymnosperm. An angiosperm is a seed-bearing plant whose seeds are borne in a mature ovary (fruit) and are divided into two broad classes based on the number of cotyledons or seed leaves that generally store or absorb food. A gymnosperm is a seed-bearing plant with seeds not enclosed in an ovary. In view of the above, the skilled person understands that the invention can be practiced, for example, with a monocotyledonous or dicotyledonous angiosperm or gymnosperm as desired.

In one embodiment, the invention provides a transgenic woody plant that is characterized by suppressed flowering. A transgenic plant of the invention can be, for example, a perennial woody plant such as a tree or shrub. For example, dicot trees such as alder, ash, basswood, beech, birch, cherry, cottonwood, elm, hickory, locust, maple, red and white oak, persimmon, sycamore, walnut, and poplar can be modified as disclosed herein to produce transgenic varieties in which flowering is suppressed. In addition, conifer woods, for example, cedar; Douglas fir; hemlock; loblolly, ponderosa, slash, sugar and western white pines; redwood; and spruce trees can be modified to produce transgenic varieties in which flowering is suppressed. The skilled person understands that the invention can be practiced with these or other shrubs or trees, especially trees useful for producing lumber, pulp or paper (Whetten and Sederoff, Forest Ecology and Management 43:301-316 (1991), which is incorporated herein by reference).

The present invention further provides tissues derived from a transgenic plant of the invention. Such tissues are derived from a transgenic plant that is characterized by suppressed flowering and that contains a nucleic acid molecule including a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product.

As used herein, the term "tissue" means an aggregate of plant cells and intercellular material organized into a structural and functional unit. A particularly useful tissue of the

invention is a tissue that can be vegetatively or non-vegetatively propagated such that the transgenic plant from which the tissue was derived is reproduced. A tissue of the invention can be, for example, a leaf, root, stem or part thereof.

The present invention also provides an isolated nucleic acid molecule including an  
5 *AGL2*, *AGL4* or *AGL9* or *API* regulatory element, which confers selective expression upon an operatively linked nucleotide sequence in one or more floral organs of a plant. The isolated nucleic acid molecule can further include, if desired, an operatively linked nucleotide sequence encoding a cytotoxic gene product. The encoded cytotoxic gene product can be, for example, diphtheria toxin A chain, RNase T1, Barnase RNase, ricin toxin A chain, or the  
10 herpes simplex virus thymidine kinase gene product.

The *Arabidopsis AGL2* promoter (SEQ ID NO:1) is shown in Figure 1. An *AGL2* regulatory element, such as a 5' regulatory element or intronic regulatory element, can confer selective expression in one or more floral organs such as carpels and stamens and, thus, is a floral organ selective regulatory element as defined herein. An isolated *AGL2* floral organ  
15 selective regulatory element can have, for example, at least fifteen contiguous nucleotides of the *Arabidopsis AGL2* sequence SEQ ID NO:1. Such an isolated *AGL2* floral organ selective regulatory element can have, for example, at least 16, 18, 20, 25, 30, 40, 50, 100 or 500 contiguous nucleotides of SEQ ID NO:1 and is characterized, in part, by the ability to confer floral organ selective expression upon an operatively linked nucleotide sequence (see  
20 Example I).

The *Arabidopsis AGL4* promoter (SEQ ID NO:2) is shown in Figure 2. An *AGL4* regulatory element confers selective expression in one or more floral organs without conferring expression in vegetative tissues and, thus, is a floral organ selective regulatory element as defined herein. An isolated *AGL4* floral organ selective regulatory element can  
25 have, for example, at least fifteen contiguous nucleotides of the *Arabidopsis AGL4* sequence SEQ ID NO:2. Such an isolated *AGL4* floral organ selective regulatory element can have, for example, at least 16, 18, 20, 25, 30, 40, 50, 100 or 500 contiguous nucleotides of SEQ ID NO:2 and is characterized, in part, by the ability to confer floral organ selective expression upon an operatively linked nucleotide sequence (see Example II).

The *Arabidopsis AGL9* promoter (SEQ ID NO:3) is shown in Figure 3. An *AGL9* regulatory element, such as a 5' regulatory element or intronic regulatory element, can confer selective expression in one or more floral organs, specifically in petals, stamens and carpels, and, thus, is a floral organ selective regulatory element as defined herein. An isolated *AGL9*  
30

floral organ selective regulatory element can have, for example, at least fifteen contiguous nucleotides of the *Arabidopsis AGL9* sequence SEQ ID NO:3. Such an isolated *AGL9* floral organ selective regulatory element can have, for example, at least 16, 18, 20, 25, 30, 40, 50, 100 or 500 contiguous nucleotides of SEQ ID NO:3 and is characterized, in part, by the ability to confer floral organ selective expression upon an operatively linked nucleotide sequence (see Example III).

The *Arabidopsis API* promoter (SEQ ID NO:10) is shown in Figure 6. An *API* regulatory element, such as a 5' regulatory element or intronic regulatory element, can confer selective expression in one or more floral organs, specifically in petals, stamens and carpels, and, thus, is a floral organ selective regulatory element as defined herein. An isolated *API* floral organ selective regulatory element can have, for example, at least fifteen contiguous nucleotides of the *Arabidopsis API* sequence SEQ ID NO:10. Such an isolated *API* floral organ selective regulatory element can have, for example, at least 16, 18, 20, 25, 30, 40, 50, 100 or 500 contiguous nucleotides of SEQ ID NO:10 and is characterized, in part, by the ability to confer floral organ selective expression upon an operatively linked nucleotide sequence (see Example IV).

As used herein, the term "substantially the nucleotide sequence," when used in reference to an *AGL2*, *AGL4* or *AGL9* or *API* regulatory element, means a nucleotide sequence having an identical sequence, or a nucleotide sequence having a similar, non-identical sequence that is considered to be a functionally equivalent sequence by those skilled in the art. For example, a floral organ selective regulatory element that is an *AGL2* regulatory element can have, for example, a nucleotide sequence identical to the sequence of the *Arabidopsis AGL2* promoter (SEQ ID NO:1) shown in Figure 1, or a similar, non-identical sequence that is functionally equivalent. A floral organ selective regulatory element can have, for example, one or more modifications such as nucleotide additions, deletions or substitutions relative to the *AGL2* promoter sequence shown in Figure 1, provided that the modified nucleotide sequence retains substantially the ability to confer selective expression in one or more floral organs upon an operatively linked nucleotide sequence, such as a nucleotide sequence encoding a cytotoxic gene product.

It is understood that limited modifications can be made without destroying the biological function of an *AGL2*, *AGL4* or *AGL9* or *API* regulatory element and that such limited modifications can result in floral organ selective regulatory elements that have substantially equivalent or enhanced function as compared to a wild type *AGL2*, *AGL4* or

AGL9 or AP1 regulatory element. These modifications can be deliberate, as through site-directed mutagenesis, or can be accidental such as through mutation in hosts harboring the regulatory element. All such modified nucleotide sequences are included in the definition of a floral organ selective regulatory element as long as the ability to confer selective expression in one or more floral organs is substantially retained.

A floral organ selective regulatory element can be derived from a gene that is an ortholog of *Arabidopsis* AGL2, AGL4 or AGL9 or AP1 and that is selectively expressed in one or more floral organs of the orthologous plant. An AGL2, AGL4 or AGL9 or AP1 floral organ selective regulatory element can be derived, for example, from an AGL2, AGL4 or AGL9 or AP1 ortholog such as a Eucalyptus, pine or spruce ortholog.

Floral organ selective regulatory elements also can be derived from a variety of other genes that are selectively expressed in one or more floral organs of a plant and can be identified and isolated using routine methodology. Differential screening strategies using, for example, RNA prepared from a floral organ and RNA prepared from non-floral material such as leaf or root tissue can be used to isolate cDNAs selectively expressed in cells of one or more floral organs; subsequently, the corresponding genes are isolated using the cDNA sequence as a probe.

Enhancer trap or gene trap strategies also can be used to identify and isolate a floral organ selective regulatory element (Sundaresan, et al., Genes Dev. 9, 1797-1810 (1995); Koncz et al., Proc. Natl. Acad. Sci. USA 86:8467-8471 (1989); Kertbundit et al., Proc. Natl. Acad. Sci. USA 88:5212-5216 (1991); Topping et al., Development 112:1009-1019 (1991), each of which is incorporated herein by reference). Enhancer trap elements include a reporter gene such as GUS with a weak or minimal promoter, while gene trap elements lack a promoter sequence, relying on transcription from a flanking chromosomal gene for reporter gene expression. Transposable elements included in the constructs mediate fusions to endogenous loci; constructs selectively expressed in one or more floral organs are identified by their pattern of expression. With the inserted element as a tag, the flanking floral organ selective regulatory element is cloned using, for example, inverse polymerase chain reaction methodology (see, for example, Aarts et al., Nature 363:715-717 (1993); see, also, Ochman et al., "Amplification of Flanking Sequences by Inverse PCR," in Innis et al. (Ed.), PCR Protocols, San Diego: Academic Press, Inc. (1990)). The Ac/Ds transposition system of Sundaresan, et al., Genes Dev. 9, 1797-1810 (1995), can be particularly useful in identifying and isolating a floral organ selective regulatory element useful in the invention.

Floral organ selective regulatory elements also can be isolated by inserting a library of random genomic DNA fragments in front of a promoterless reporter gene and screening transgenic plants transformed with the library for floral organ selective reporter gene expression. The promoterless vector pROA97, which contains the *npt* gene and the GUS gene each under the control of the minimal 35S promoter, can be useful for such screening. The genomic library can be, for example, Sau3A fragments of *Arabidopsis thaliana* genomic DNA or genomic DNA from, for example, Eucalyptus, pine or spruce (Ott et al., Mol. Gen. Genet. 223:169-179 (1990); Claes et al., The Plant Journal 1:15-26 (1991), each of which is incorporated herein by reference).

An active fragment of an *AGL2*, *AGL4* or *AGL9* or *API* promoter, which contains a floral organ selective regulatory element, can be identified by routine techniques, for example, using a reporter gene and *in situ* expression analysis. The GUS and firefly luciferase reporter genes are particularly useful for *in situ* localization of plant gene expression (Jefferson et al., EMBO J. 6:3901 (1987); Ow et al., Science 334:856 (1986), each of which is incorporated herein by reference), and promoterless vectors containing the GUS expression cassette are commercially available, for example, from Clontech (Palo Alto, CA). To identify an active fragment containing a floral organ selective regulatory element such as an *AGL2*, *AGL4* or *AGL9* or *API* regulatory element, one or more nucleotide portions of an *AGL2*, *AGL4* or *AGL9* or *API* gene can be generated using enzymatic or PCR-based methodology (Glick and Thompson (eds.), Methods in Plant Molecular Biology and Biotechnology, Boca Raton, FL: CRC Press (1993); Innis et al. (Ed.), PCR Protocols, San Diego: Academic Press, Inc. (1990)); the resulting segments are fused to a reporter gene such as GUS and analyzed as described above.

The present invention also provides a kit for producing a transgenic plant characterized by suppressed flowering. A kit of the invention comprises packaging containing a plant expression vector having a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product. The plant expression vector can include, if desired, a nucleotide sequence encoding a selectable marker or reporter gene, along with instructions to employ the vector in accord with the present method.

The term "plant expression vector," as used herein, is a self-replicating nucleic acid molecule that provides a means to transfer an exogenous nucleic acid molecule into a plant host cell and to express the molecule therein. Plant expression vectors encompass vectors

suitable for *Agrobacterium*-mediated transformation, including binary and cointegrating vectors, as well as vectors for physical transformation.

Plant expression vectors can be used for transient expression of the exogenous nucleic acid molecule, or can integrate and stably express the exogenous sequence. One skilled in the art understands that a plant expression vector can contain all the functions needed for transfer and expression of an exogenous nucleic acid molecule; alternatively, one or more functions can be supplied in *trans* as in a binary vector system for *Agrobacterium*-mediated transformation.

In addition to a floral organ selective regulatory element and a nucleotide sequence encoding a cytotoxic gene product, a plant expression vector of the invention can contain, if desired, additional elements. A binary vector for *Agrobacterium*-mediated transformation contains one or both T-DNA border repeats and can also contain, for example, one or more of the following: a broad host range replicon, an *ori T* for efficient transfer from *E. coli* to *Agrobacterium*, a bacterial selectable marker such as ampicillin and a polylinker containing multiple cloning sites.

A plant expression vector for physical transformation can have, if desired, a plant selectable marker or a reporter gene or both, in addition to a floral organ selective regulatory element in vectors such as pBR322, pUC, pGEM and M13, which are commercially available, for example, from Pharmacia (Piscataway, NJ) or Promega (Madison, WI).

A selectable marker gene or a reporter gene can facilitate the identification and selection of transformed plants, or plant cells. Both selectable marker and reporter genes may be flanked with appropriate regulatory sequences to enable expression in plants. Useful selectable markers are well known in the art and include, for example, antibiotic and herbicide resistance genes. Specific examples of such genes are disclosed in Weising, K., et al., *Ann. Rev. Genet.*, 22, 421-478 (1988). Selectable marker genes includes the hygromycin B phosphotransferase coding sequence, which confers resistance to hygromycin B; the aminoglycoside phosphotransferase gene of transposon Tn5 (AphII), which encodes resistance to the antibiotics kanamycin, neomycin and G418; and genes which code for resistance or tolerance to glyphosate, 1,2-dichloropropionic acid methotrexate, imidazolinones, sulfonylureas, bromoxynil, phosphonothricin and the like.

Reporter genes which encode for easily assayable marker proteins are well known in the art. IN general, a reporter gene is a gene which ins not present in or expressed by the recipient organism or tissue and which encodes a protein whose expression is manifested by

some easily detectable property, e.g., phenotypic change or enzymatic activity. Examples of such gene are provided in Weising, et al., Ann. Rev. Genet., 22, 421-478 (1988).

In plant expression vectors for physical transformation of a plant, the T-DNA borders or the *ori* T region can optionally be included but provide no advantage.

5 Also provided by the present invention is a method of producing a transgenic plant characterized by suppressed flowering. The method includes the step of introducing into a plant an exogenous nucleic acid molecule containing a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product, where flowering is suppressed due to selective expression of the exogenous nucleic acid molecule and where the floral organ selective regulatory element is an *AGL2* regulatory element, an  
10 *AGL4* regulatory element or an *AGL9* regulatory element or an *API* regulatory element.

Methods for producing the desired recombinant nucleic acid molecule under control of an *AGL2*, *AGL4* or *AGL9* or *API* floral organ selective regulatory element and for producing a transgenic plant of the invention are well known in the art (see, generally, Sambrook et al.  
15 (eds.) Molecular Cloning: A Laboratory Manual (Second Edition, Plainview, NY: Cold Spring Harbor Laboratory Press (1989); Glick and Thompson, *supra*, 1993).

An exogenous nucleic acid molecule can be introduced into a plant using a variety of transformation methodologies including *Agrobacterium*-mediated transformation and direct gene transfer methods such as electroporation and microprojectile-mediated transformation  
20 (see, generally, Wang et al. (eds), Transformation of Plants and Soil Microorganisms, Cambridge, UK: University Press (1995), which is incorporated herein by reference).

Transformation methods based upon the soil bacterium *Agrobacterium tumefaciens* are particularly useful for introducing an exogenous nucleic acid molecule into a plant. The wild type form of *Agrobacterium* contains a Ti (tumor-inducing) plasmid that directs production of  
25 tumorigenic crown gall growth on host plants. Transfer of the tumor-inducing T-DNA region of the Ti plasmid to a plant genome requires the Ti plasmid-encoded virulence genes as well as T-DNA borders, which are a set of direct DNA repeats that delineate the region to be transferred. An *Agrobacterium*-based vector is a modified form of a Ti plasmid, in which the tumor inducing functions are replaced by the nucleic acid sequence of interest to be  
30 introduced into the plant host.

*Agrobacterium*-mediated transformation generally employs cointegrate vectors or, preferably, binary vector systems, in which the components of the Ti plasmid are divided between a helper vector, which resides permanently in the *Agrobacterium* host and carries the



virulence genes, and a shuttle vector, which contains the gene of interest bounded by T-DNA sequences. A variety of binary vectors are well known in the art and are commercially available, for example, from Clontech (Palo Alto, CA). Methods of coculturing *Agrobacterium* with cultured plant cells or wounded tissue such as leaf tissue, root explants, hypocotyledons, stem pieces or tubers, for example, also are well known in the art (Glick and Thompson, *supra*, 1993). Wounded cells within dicot plant tissue that have been infected by *Agrobacterium* can develop organs *de novo* when cultured under the appropriate conditions; the resulting transgenic shoots eventually give rise to transgenic plants that ectopically express a nucleic acid molecule containing a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product. *Agrobacterium* also can be used for transformation of whole plants as described in Bechtold et al., C.R. Acad. Sci. Paris, Life Sci. 316:1194-1199 (1993), which is incorporated herein by reference).

Microprojectile-mediated transformation also can be used to produce a transgenic plant containing a nucleic acid molecule including a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product. This method, as described by Lundquist et al., U.S. Pat. No. 5,554,798, which is incorporated herein by reference), relies on microprojectiles such as gold or tungsten that are coated with the desired nucleic acid molecule by precipitation with calcium chloride, spermidine or PEG. The microprojectile particles are accelerated at high speed into an angiosperm tissue using a device such as the BIOLISTIC PD-1000 (Biorad; Hercules CA).

Microprojectile-mediated delivery or "particle bombardment" is especially useful to transform plants that are difficult to transform or regenerate using other methods. Microprojectile-mediated transformation has been used, for example, to generate a variety of transgenic plant species, including cotton, tobacco, corn, hybrid poplar and papaya (see Glick and Thompson, *supra*, 1993) as well as cereal crops such as wheat, oat, barley, sorghum and rice (Duan et al., Nature Biotech. 14:494-498 (1996); Shimamoto, Curr. Opin. Biotech. 5:158-162 (1994), each of which is incorporated herein by reference). In view of the above, the skilled artisan will recognize that *Agrobacterium*-mediated or microprojectile-mediated transformation, as disclosed herein, or other methods known in the art can be used to produce a transgenic plant of the invention characterized by suppressed flowering.

Following transformation via any method, it is necessary to identify and select those plants or cells which both contain the heterologous DNA and still retain sufficient

regenerative capacity. There are two general approaches which have been found useful for accomplishing this. First, the transformed calli or plants regenerated therefrom can be screened for the presence of the heterologous DNA by various standard methods which could include assays for the expression of reporter genes or assessment of phenotypic effects of the heterologous DNA, if any. Alternatively, and preferably, when a selectable marker gene has been transmitted along with or as part of the heterologous DNA, those cells of the callus or plant which have been transformed can be identified by the use of a selective agent to detect expression of the selectable marker gene.

Selection of the putative transformants is a critical part of the successful transformation process since selection conditions must be chosen so as to allow growth and accumulation of the transformed cells or plants while simultaneously inhibiting the growth of the non-transformed cells or plants.

Selection procedures involve exposure to a toxic agent and may employ sequential changes in the concentration of the agent and multiple rounds of selection. The particular concentrations and cycle lengths are likely to need to be varied for each particular agent. A currently preferred selection procedure entails using an initial selection round at a relatively low toxic agent concentration and then later round(s) at higher concentration(s). This allows the selective agent to exert its toxic effect slowly over a longer period of time. Preferably, the concentration of the agent is initially such that about a 5-40% level of growth inhibition will occur, as determined from a growth inhibition curve. The effect may be to allow the transformed cells or plants to preferentially grow and divide while inhibiting untransformed cells or plants, but not to the extent that growth of the transformed cells or plants is prevented. Once the few individual transformed cells or plants have grown sufficiently, the tissue may be shifted to media containing a higher concentration of the toxic agent to kill essentially all untransformed cells. The shift to the higher concentration also reduces the possibility of non-transformed cells or plants habituating to the agent. The higher level is preferably in the range of about 30 to 100% growth inhibition. The length of the first selection cycle may be from about 1 to 4 weeks, preferably about 2 weeks. Later selection cycles may be from about 1 to about 12 weeks, preferably about 2 to about 10 weeks.

Putative transformants can generally be identified as viable plants. In the case of transformation of cells, putative transformants can generally be identified as proliferating sectors of tissue among a background of non-proliferating cells.

Once a putative transformant is identified, transformation can be confirmed by phenotypic and/or genotypic analysis. If a selection agent is used, an example of phenotypic analysis is to visually inspect the plants. The plants which appear to be green, growing, and healthy are compared to a control on various levels of the selective agent. Another example of phenotypic analysis is to measure the increase in fresh weight of the putative transformant as compared to a control on various levels of the selective agent. Other analyses that may be employed will depend on the function of the heterologous DNA. For example, if an enzyme or protein is encoded by the DNA, enzymatic or immunological assays specific for the particular enzyme or protein may be used. Other gene products may be assayed by using a suitable bioassay or chemical assay. Other such techniques are well known in the art and are not repeated here. The presence of the gene can also be confirmed by conventional procedures, i.e., Southern blot or polymerase chain reaction (PCR) or the like.

#### EXAMPLE I

#### AN AGL2 REGULATORY ELEMENT DIRECTS FLORAL ORGAN SELECTIVE EXPRESSION

This example shows that a fragment of the *Arabidopsis AGL2* promoter is sufficient to direct floral organ selective gene expression.

*Agrobacterium tumefaciens* strain C58 was used to transform *Arabidopsis thaliana*, ecotype Columbia. The transformation method of this example was disclosed by Bechtold et al., *C. R. Acad. Sci. Paris*, 316:1194-9 (1993)(incorporated by reference herein).

A BglII fragment of approximately 2.3 kb was isolated from the *Arabidopsis AGL2* promoter (SEQ ID NO:1) shown in Figure 1 using the BglII sites indicated at nucleotide 1 and nucleotide 1120. The fragment was subcloned into the BamHI site of pGEM3Z (Promega, Madison, WI). The resulting plasmid was restricted with SalI and SmaI and subcloned into the corresponding sites of the GUS expression vector pBI101.2 (CLONTECH, Palo Alto, CA) to create pKY18. Analysis of GUS expression in kanamycin resistant *Arabidopsis* lines transformed with pKY18 revealed floral specific GUS expression with no significant expression in tissues other than flowers.

These results indicate that the 2.3 kb *Arabidopsis AGL2* promoter fragment of SEQ ID NO:1 directs floral organ selective expression of a heterologous linked gene product.

## EXAMPLE II

AN AGL4 REGULATORY ELEMENT DIRECTS FLORAL ORGAN SELECTIVE  
EXPRESSION

This example shows that a fragment of the *Arabidopsis AGL4* promoter is sufficient to  
5 direct floral organ selective gene expression.

*Agrobacterium tumefaciens* strain C58 was used to transform *Arabidopsis thaliana*,  
ecotype Columbia. The transformation method of this example was disclosed by Bechtold et  
al., *C. R. Acad. Sci. Paris*, 316:1194-9 (1993)(incorporated by reference herein).

*AGL4* promoter fragments were isolated from the promoter sequence shown in Figure 2  
10 (SEQ ID NO:2). A 560 bp *AGL4* fragment of SEQ ID NO:2 was prepared containing the  
region from nucleotide -862 to nucleotide -303 using the HindIII site indicated at nucleotide -  
862 and an engineered BamHI site. The 560 bp fragment was subcloned into the HindIII and  
BamHI sites of pGEM3Z (Promega). A 270 bp *AGL4* fragment of SEQ ID NO:2 was  
15 prepared similarly using the indicated DraI site at nucleotide -573 and an engineered BamHI  
site at nucleotide -303 and subcloned into the HincII and BamHI sites of pGEM3Z. The 560  
bp and 270 bp fragments were subsequently cloned into the GUS expression vector pBI101.1  
(CLONTECH) to produce pSR34 and pSR35, respectively.

Plants were transformed with pSR34 and pSR35. GUS staining was observed in the  
flowers of pSR34 plants. These results demonstrate that the 560 bp fragment of the  
20 *Arabidopsis AGL4* promoter confers floral organ selective expression upon a linked gene.

## EXAMPLE III

AN AGL9 REGULATORY ELEMENT DIRECTS FLORAL ORGAN SELECTIVE  
EXPRESSION

This example shows that a fragment of the *Arabidopsis AGL9* promoter is sufficient to  
25 direct floral organ selective gene expression.

*Agrobacterium tumefaciens* strain C58 was used to transform *Arabidopsis thaliana*,  
ecotype Columbia. The transformation method of this example was disclosed by Bechtold et  
al., *C. R. Acad. Sci. Paris*, 316:1194-9 (1993)(incorporated by reference herein).

The entire 1755 bp *AGL9* promoter fragment shown in Figure 3 (SEQ ID NO:3) was  
30 cloned into the GUS expression vector pBI101.3 (CLONTECH) to produce pSP112.  
Multiple transgenic lines containing pSP112 were analyzed for GUS expression. The results

showed that GUS was expressed only in floral organs, with no expression evident in other tissues such as stem.

These results demonstrate that an *AGL9* promoter is a floral organ selective regulatory element that can confer floral organ selective expression upon an operatively linked encoded gene such as GUS.

#### EXAMPLE IV

#### AN *API* REGULATORY ELEMENT DIRECTS FLORAL ORGAN SELECTIVE EXPRESSION

This example shows that a fragment of the *Arabidopsis API* promoter is sufficient to direct floral selective gene expression.

*Agrobacterium tumefaciens* strain C58 was used to transform *Arabidopsis thaliana*, ecotype Columbia. The transformation method of this example was disclosed by Bechtold et al., C. R. Acad. Sci. Paris, 316:1194-9 (1993)(incorporated by reference herein).

The entire 1.7 kb *API* promoter shown in Figure 6 (SEQ ID NO: 10) plus the entire coding region of *API* including introns was cloned into the GUS expression vector pBI101.2 to produce the POP10 construct (Figure 7). The construct was first made by PCR amplification from intron 3 to the end of *API* gene in exon 8 (right before stop codon) using KY65 plasmid containing *API* genomic region as template. The HindIII site was added to the forward primer AP1HIN and BamHI site was added to reverse primer AP1BAM to aid cloning. The 1.7 kb amplified fragment was cloned into plasmid pBI101.2 using HindIII and BamHI sites giving construct POP9. The 3.6 kb HindIII / XbaI fragment was isolated from KY65 plasmid and cloned into POP9 construct giving POP10 construct.

Multiple transgenic lines containing the POP10 construct were analyzed for GUS expression. The results showed the GUS was expressed specifically in the young flower primordium (See Figure 9) as soon as it arises on the flanks of the shoot meristem. No GUS staining was seen in the shoot meristem, the stem, leaves, roots, or any part of the plant other than in flowers.

All journal articles, references, and patent citations provided above, in parentheses or otherwise, whether previously stated or not, are incorporated herein by reference.

It should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims.

REFERENCES CITED

The references listed below are incorporated by reference herein.

- Aarts et al., Nature 363:715-717 (1993).
- Angenent et al., Plant Cell 4:983-993 (1992).
- 5 Bechtold et al., C.R. Acad. Sci. Paris, Life Sci. 316:1194-1199 (1993).
- Bowman et al., Devel. 112:1-20 (1991).
- Bowman et al., The Plant Cell 1:37-52 (1989).
- Claes et al., The Plant Journal 1:15-26 (1991).
- Collier, Bacteriol. Rev. 39:54-85 (1975).
- 10 Davies et al., EMBO J. 15:4330-4343 (1996).
- Duan et al., Nature Biotech. 14:494-498 (1996).
- Glick and Thompson (eds.), Methods in Plant Molecular Biology and Biotechnology, Boca Raton, FL: CRC Press (1993).
- Greenfield et al., Proc. Natl. Acad. Sci., USA 80:6853-6857 (1983).
- 15 Higuchi, "Recombinant PCR" in Innis et al. (Ed.), PCR Protocols, San Diego: Academic Press, Inc. (1990).
- Huang et al., The Plant Cell 8:81-94 (1996).
- Huijser et al., EMBO J. 33:1239-1249; 1992.
- Innis et al. (Ed.), PCR Protocols, San Diego: Academic Press, Inc. (1990).
- 20 Jefferson et al., EMBO J. 6:3901 (1987).
- Kang and An, Mol. Cells 7:45-51 (1997).
- Kertbundit et al., Proc. Natl. Acad. Sci. USA 88:5212-5216 (1991).
- Koncz et al., Proc. Natl. Acad. Sci. USA 86:8467-8471 (1989).
- Liu and Podila, Plant Phys. 113:665 (1997).
- 25 Ma et al., Genes Devel. 5:484-495 (1991).
- Mandel and Yanofsky, The Plant Cell 7:1763-1771 (1995).

05869582-022002

- Mandel, Nature 360:273-277 (1992).
- Moffat et al., Development 114:681-687 (1992).
- Mouradov et al., Plant Physiol. 117:55-62 (1998).
- Münster et al., Proc. Natl. Acad. Sci., USA 94:2145-2420 (1997).
- 5 Norman et al., Cell 55:989-1003 (1988).
- Ochman et al., "Amplification of Flanking Sequences by Inverse PCR," in Innis et al., (Ed.), PCR Protocols, San Diego: Academic Press, Inc. (1990).
- Olsnes and Pihl, Molecular Action of Toxins and Viruses, pages 51-105, Amsterdam: Elsevier Biomedical Press (1982).
- 10 Ott et al., Mol. Gen. Genet. 223:169-179 (1990).
- Ow et al., Science 334:856 (1986).
- Palmiter et al., Cell 50:435-443 (1987).
- Passmore et al., J. Mol. Biol. 204:593-606 (1988).
- Pnueli et al., Plant J. 1:255-266 (1991).
- 15 Pnueli et al., The Plant Cell 6:163-173 (1994).
- Purugganan et al., Genetics 140:345-356 (1995).
- Riechmann and Meyerowitz, Biol. Chem. 378:1079-1101 (1997).
- Sambrook et al. (eds.) Molecular Cloning: A Laboratory Manual (Second Edition, Plainview, NY: Cold Spring Harbor Laboratory Press (1989).
- 20 Schmidt et al., The Plant Cell 5:729-737 (1993).
- Shimamoto, Curr. Opin. Biotech. 5:158-162 (1994).
- Sundaresan, et al., Genes Dev. 9, 1797-1810 (1995).
- Tandre et al., Plant Mol. Biol. 27:69-78 (1995).
- Topping et al., Development 112:1009-1019 (1991).
- 25 Thorsness and Nasrallah, Methods in Cell Biology 50:439-448 (1995).
- Wang et al. (eds), Transformation of Plants and Soil Microorganisms, Cambridge, UK: University Press (1995).

09865582, 022802

Weigel and Meyerowitz, Cell 78:203-209 (1994).

Weising, K., et al., Ann. Rev. Genet., 22, 421-478 (1988).

Whetten and Sederoff, Forest Ecology and Management 43:301-316 (1991).

Yamaizumi et al., Cell 15:245-250 (1978).

5 Yanofsky, Annual Rev. Plant Physiol. Mol. Biol. 46:167-188 (1995).

Yanofsky et al., Nature 346:35-39 (1990).

09065582 022802



What is claimed is:

1. A transgenic plant characterized by suppressed flowering, comprising a nucleic acid molecule comprising a floral organ selective regulatory element, operatively linked to a nucleotide sequence encoding a cytotoxic gene product, wherein said nucleic acid molecule is heritable by progeny thereof.
2. The transgenic plant of claim 1, wherein said floral organ selective regulatory element is selected from the group consisting of an *AGL2* regulatory element, *AGL4* regulatory element, *AGL9* regulatory element, and an *AP1* regulatory element.
3. The transgenic plant of claim 1, wherein said cytotoxic gene product is selected from the group consisting of diphtheria toxic A chain, RNase T1, Barnase Rnase, ricin toxin A chain, and herpes simplex virus thymidine kinase (tk) gene.
4. The transgenic plant of claim 2, wherein said *AGL2* regulatory element has substantially the nucleotide sequence of *Arabidopsis AGL2* promoter SEQ ID NO:1, or an active fragment thereof.
5. The transgenic plant of claim 2, wherein said *AGL4* regulatory element has substantially the nucleotide sequence of *Arabidopsis AGL4* promoter SEQ ID NO:2, or an active fragment thereof.
6. The transgenic plant of claim 2, wherein said *AGL9* regulatory element has substantially the nucleotide sequence of *Arabidopsis AGL9* promoter SEQ ID NO:3, or an active fragment thereof.
7. The transgenic plant of claim 2, wherein said *AP1* regulatory element has substantially the nucleotide sequence of *Arabidopsis AP1* promoter SEQ ID NO:10, or an active fragment thereof.
8. A tissue derived from the transgenic plant of any of claims 1 to 7.

9. The tissue of claim 8, which is capable of non-vegetative propagation.
10. The tissue of claim 8, which is capable of vegetative propagation.
11. The plant of claim 1, wherein said plant is a woody plant.
12. The plant of claim 11, wherein said plant is a tree.
- 5 13. A method of producing a transgenic plant characterized by suppressed flowering, comprising introducing into a plant an exogenous nucleic acid molecule comprising a floral organ selective regulatory element, wherein said regulatory element is operatively linked to a nucleotide sequence encoding a cytotoxic gene product, whereby flowering is suppressed due to selective expression of said exogenous nucleic acid molecule in said floral organ, and
- 10 wherein said nucleic acid molecule is heritable by progeny thereof.
14. The method of claim 13, wherein said floral organ selective regulatory element is selected from the group consisting of an *AGL2* regulatory element, *AGL4* regulatory element, *AGL9* regulatory element, and an *API* regulatory element.
- 15 15. The method of claim 14, wherein said *AGL2* regulatory element has substantially the nucleotide sequence of *Arabidopsis AGL2* promoter SEQ ID NO:1, or an active fragment thereof.
16. The method of claim 14, wherein said *AGL4* regulatory element has substantially the nucleotide sequence of *Arabidopsis AGL4* promoter SEQ ID NO:2, or an active fragment thereof.
- 20 17. The method of claim 14, wherein said *AGL9* regulatory element has substantially the nucleotide sequence of *Arabidopsis AGL9* promoter SEQ ID NO:3, or an active fragment thereof.

09869582-02280

18. The method of claim 14, wherein said *AP1* regulatory element has substantially the nucleotide sequence of *Arabidopsis AP1* promoter SEQ ID NO:10, or an active fragment thereof.

5 19. The method of claim 13, wherein said cytotoxic gene product is selected from the group consisting of diphtheria toxic A chain, RNase T1, Barnase Rnase, ricin toxin A chain, and herpes simplex virus thymidine kinase (tk) gene.

20. The method of claim 13, wherein the nucleic acid molecule is introduced into the plant by *Agrobacterium*-mediated transformation.

10 21. The method of claim 20, wherein *Agrobacterium tumefaciens* is used to introduce the nucleic acid molecule into the plant.

22. The method of claim 20, wherein *Agrobacterium rhizogenes* is used to introduce the nucleic acid molecule into the plant.

15 23. The transgenic plant of claim 1, wherein said plant is obtainable by a process comprising the steps of (i) introducing into a plant an exogenous nucleic acid molecule comprising a floral organ selective regulatory element, wherein said regulatory element is operatively linked to a nucleotide sequence encoding a cytotoxic gene product; (ii) identifying or selecting a population of plants whose flowering is suppressed; (iii) generating a progeny transgenic plant therefrom.

20 24. An isolated nucleic acid molecule, comprising a floral organ selective regulatory element, operatively linked to a nucleotide sequence encoding a cytotoxic gene product.

25 25. The isolated nucleic acid molecule of claim 24, wherein said regulatory element is selected from the group consisting of an *AGL2* regulatory element, *AGL4* regulatory element, *AGL9* regulatory element, and an *AP1* regulatory element.

26. The isolated nucleic acid molecule of claim 25, comprising at least fifteen contiguous nucleotides of *Arabidopsis AGL2* promoter SEQ ID NO:1.

27. The isolated nucleic acid molecule of claim 25, comprising at least fifteen contiguous nucleotides of *Arabidopsis AGL4* promoter SEQ ID NO:2.
28. The isolated nucleic acid molecule of claim 25, comprising at least fifteen contiguous nucleotides of *Arabidopsis AGL9* promoter SEQ ID NO:3.
- 5 29. The isolated nucleic acid molecule of claim 25, comprising at least fifteen contiguous nucleotides of *Arabidopsis AP1* promoter SEQ ID NO:10.
30. The isolated nucleic acid molecule of claim 24, wherein said cytotoxic gene product is selected from the group consisting of diphtheria toxic A chain, RNase T1, Barnase Rnase, ricin toxin A chain, and herpes simplex virus thymidine kinase (tk) gene.
- 10 31. A kit for producing a transgenic plant characterized by suppressed flowering, comprising packaging containing a plant expression vector comprising a floral organ selective regulatory element operatively linked to a nucleotide sequence encoding a cytotoxic gene product, and instructions for transforming a susceptible plant with said vector.
32. The kit of claim 31, wherein said regulatory element is selected from the group  
15 consisting of an *AGL2* regulatory element, *AGL4* regulatory element, *AGL9* regulatory element, and an *AP1* regulatory element.
33. The kit of claim 31, wherein said cytotoxic gene product is selected from the group consisting of diphtheria toxic A chain, RNase T1, Barnase Rnase, ricin toxin A chain, and herpes simplex virus thymidine kinase (tk) gene.

09369582-022802

Sequence Range: 1 to 4512

AGATCTCTAT GAAAAATGGC AAAATCAACA ATAATCCCTT GGCTATATGG TGGTATTCTCT  
TCTAGAGATA CTTTTTACCG TTTTAGTTGT TATTAGGGAA CCGATATAACC ACCATAAAGA

50

GTTAAAAGTG ACTTATGGGT AGATTTTTTA GCTTCATAGA TTCTTTGTCG AAAAAAATT  
CAATTTTCAC TGAATACCCA TCTAAAAAAT CGAAGTATCT AAGAAACAGC TTTTTTTTAA

100

ACTTTGTACA TTTTAGTGGA GTTATTTAAA TTTCCCAATT GAACAAAACC ATATATTGAT  
TGAAACATGT AAAATCACCT CAATAAATTT AAAGGGTTAA CTGTGTTTGG TATATAACTA

150

GAAATTGCGA AATGCAATCC AAAAATAAAT ATGTTCCACT CTTTTGGTTA GCTTTTAACT  
CTTAAAGCGT TTACGTTAGG TTTTATTTTA TACAAGGTGA GAAAACCAAT CGAAAATTGA

200

AAACATGCGT TTT----- TTCCAGCTAG TACGAGTCTC TATATATAAA CTTTCTTAAT  
TTGTACGCA AAA----- AAGGTCGATC ATGCTCAGAG ATATATATTT GAAAGAATTA

250 300

ATCGCTAACA ATTTACTTCA AGTTTGTAAT GTGATAAGTG AAAGACCGTA TATACATACA  
TAGCGATTGT TAAATGAAGT TCAAACATTA CACTATTCAC TTTCTGGCAT ATATGTATGT

350

CATGTTAATC AACTGATAAC CTTTGTGCCT CGTGTGTCTA GTTACTAGTC AACCATCAAA  
GTACAATTAG TTGACTATTG GAAACACGGA GCACACAGAT CAATGATCAG TTGGTAGTTT

400

CGTGCATGAT GCTGTTTTTC TTAGAGTACT ATTGTTGTGT TATATATAAC TAAACATAAA  
GCACGTAATA CGACAAAAAG AATCTCATGA TAACAACACA ATATATATTG ATTTGTATTT

450

CAATTTGCTA TTATGATATA AACATAGAAT TTTCAAGCAA TGATATGTTT AGATGTTTTG  
GTTAAACGAT AATACTATAT TTGTATCTTA AAAGTTGCTT ACTATACAAA TCTACAAAAC

500

TATAAATATT CCATAAATAG TAGACACCCA TATATACACA AACATGAATT CTACCTGAGG  
ATATTTATAA GGTATTTATC ATCTGTGGGT ATATATGTGT TTGTACTTAA GATGGACTCC

550 600

AGAAACACAT AGATGTTCAA ATTAAATAAT AACCCATAA TGAAAACCTT AAAGTAAGTA  
TCTTTGTGTA TCTACAAGTT TAATTTATTA TTGGGATATT ACTTTTGAGA TTTTCATTCAT

650

ATACGAAATA AAAATTTATC CTTTAAATAA CATATAACAT ATATATCAAC TTTAATTGGT  
TATGCTTTAT TTTTAAATAG GAAATTTATT GTATATTGTA TATATAGTTG AAATTAACCA

700

AATTGTATCA CAAGAGCCAA TTATTTGGTG ACTGTATCAC ACGTGCTTAA AGAGAGCGTG  
TTAACATAGT GTTCTCGGTT AATAAACCAC TGACATAGTG TGCACGAATT TCTCTCGCAC

750

GGAATGAAAG TAAAGAAGAA TAAAGAAGCA GAGAGATGGG CTAGAAATGA GAAACACAC  
CCTTACTTTC ATTTCTTCTT ATTTCTTCGT CTCTCTACCC GATCTTTACT CTTTGTGTG

800

CAAACCCTAA CCTCACCTC ACACATTTCT TATCTTTTGC TCTCAATAGA TTCCATTGAT  
GTTTGGGATT GGAGTGGGAG TGTGTAAAGA ATAGAAAACG AGAGTTATCT AAGGTAACCTA

850 900

Fig. 1a

950  
TCAAAACAAA ATTTTCATTA AGATTTTACA ACCTCCACAC ACTTCCAAAC ACAATTAAAG  
AGTTTTGTTT TAAAAGTAAT TCTAAAGTGT TGGAGGTGTG TGAAGGTTTG TGTTAATTTT

1000  
AGAGGAAAAA GAATCAATAA CCCTATAAAT AAAAAATCAG ACAAACAGAA GTTTCCTCTT  
TCTCCTTTTT CTTAGTTATT GGGATATTTA TTTTITAGTC TGTTTGCTT CAAAGGAGAA

1050  
CTTCTTCCTT AAGCTAGTAC CTTTGTCTT TGAAATTAGG GTTAATTTCT TTTTCCAAA  
GAAGAAGGAA TTCGATCATG GAAAACAAGA ACTTTAATCC CAATTAAAGA AAAAAGGTTT

1100  
TACCATCAAT TCTCCAGACC ATAAAACTC AAAAAGATCA GATCTTTCCT CTGAAAAAGA  
ATGGTAGTTA AGAGGTCTGG TATTTTTGAG TTTTCTAGT CTAGAAAGGA GACTTTTTCT

1150 1200  
GATACCCAAC TTATGTTTTT GTGTGTCTGT ATATAGATAA ACATTACATA CCCATATTTG  
CTATGGGTTG AATACAAAAA CACACAGACA TATATCTATT TGTAATGTAT GGTATAAAG

1250  
TGTATAGACA TAAAAAGTGG AAATTAAGGT AACAAAAAGA AATGGGAAGA GGAAGAGTAG  
ACATATCTGT ATTTTTCACC TTTAATTCCA TTGTTTTTCT TTACCCCTTCT CTTTCTCATC

1300  
AGCTGAAGAG GATAGAGAAC AAAATCAACA GACAAGTAAC GTTGCAGGAG CGTAGGAACG  
TCGACTTCTC CTATCTCTTG TTTTAGTTGT CTGTTTATTG CAAACGTTTC GCATCCTTGC

1350  
GTTTGTGAA GAAAGCTTAT GAATTGTCTG TTCTCTGTGA TGCTGAAGTT GCTCTCATCA  
CAAACAACTT CTTTCGAATA CTTAACAGAC AAGAGACACT ACGACTTCAA CGAGAGTAGT

1400  
TCTTCTCAA CCGTGGAAAG CTCTATGAGT TTTGCAGCTC CTCAAAGTAA ACAACTCTCT  
AGAAGAGGTT GGCACCTTTC GAGATACTCA AAACGTCGAG GAGTTTCATT TGTGAGAGA

1450 1500  
CACTCTTTAT CAGTTTCTTG ATTGAGTTTT TGCTAGATCT GAGCTTAGAT CTTTGTCTCA  
GTGAGAAATA GTCAAAGAAC TAACTCAAAA ACGATCTAGA CTCGAATCTA GAAACAGAGT

1550  
AGGACTTGTT ATATATAGAT CACACGATCT TGATTCTAC GAAGTTGAGT TAATTAGATT  
TCCTGAACAA TATATATCTA GTGTGCTAGA ACTAAAGATG CTTCAACTCA ATTAATCTAA

1600  
TCTTGATTTT ATTTTCTAGG GTTTTTTTCC AATTCTTGAA ATTTAAGATC TGGTTTTTTT  
AGAAGTAAAG TAAAAGATCC CAAAAAAGG TTAAGAACTT TAAATTCTAG ACCAAAAAAA

1650  
GTTGTCAATG ATTTAGAACT GTGAATTTTG TAATCGAATA GATTCCAAAT CCTGATATGC  
CAACAGTTAC TAAATCTTGA CACTTAAAC ATTAGCTTAT CTAAGGTTTA GGACTATACG

1700  
AATCTGAAAA GTTTTATATA ATTAATATAT GTCTGTGTGA TTGGAACTT AAAAGTTGGA  
TTAGACTTTT CAAAATATAT TAATTATATA CAGACACACT AACCTTTGAA TTTTCAACCT

1750 1800  
ATCACAGATT TCTATGAAAA TTACAAGTAT CCAACGTAGA ATTGATAATA TATGGTTACA  
TAGTGTCTAA AGATACTTTT AATGTTTATA GGTGTCATCT TAACTATTAT ATACCAATGT

1850  
TGCATTAACC ATTTGTTAGT TCATCATACT TTATGGTGGT TAAAACTTCA AACGCGTGTA

Fig. 1b

ACGTAATTGG TAAACAATCA AGTAGTATGA AATACCACCA ATTTTGAAGT TTGCGCACAT

1900

TATCTATGAA GGCAAAGATT GTTTGTTTTT TCTTAAAAAC AATGTTTAAT AGATTTTAA  
ATAGATACTT CCGTTTCTAA CAAACAAAA AGAATTTTGT TTACAAATTA TCTAAAAAT

1950

TTATATGTTA AAATAGTTTT GCTTACATGC ATTCAAGAAA ATATAGCGAT TAATTCCTTT  
AATATACAAT TTTATCAAAA CGAATGTACG TAAGTCTTTT TATATCGCTA ATTAAGGAAA

2000

TTTCAAATCA CAATTTGTGA ATCAAACGAA AACGTAAGAT ATTGCTTGCA AATGATAGGA  
AAAGTTTAGT GTTAAACACT TAGTTTGCTT TTGCATTCTA TAACGAACGT TTACTATCCT

2050

TTGAACATTT GATATTTGTA AATATAAATA CGAAACTTTA CGTTTGAAAG TTGAAACAAT  
AACTTGATAA CTATAAACAT TTATATTTAT GCTTTGAAAT GCAAACCTTC AACTTTGTTA

2100

2150

CAAATCCAAA TCAACTCGTA TATAATCAGA TAAATAATGG AAACAATCTT CAATTTTGAT  
GTTTAGGTTT AGTTGAGCAT ATATTAGTCT ATTTATTACC TTTGTTAGAA GTTAAACTA

2200

GGAAGAATAC TTTAAACTT GAAGAGCTTT TTTTTTTTAT GGTGATTTAT AGGTTTAGAT  
CCTTCTTATG AAATTTTGAA CTCTCGAAA AAAAAAATA CCACTAAATA TCCAAATCTA

2250

CTCCAAAGTC AAGTATGATC TTTTTAATAA ACTCTTATTC TCTCTTTTGT AGTTATTTTC  
GAGGTTTCAG TTCATACTAG AAAAATTATT TGAGAATAAG AGAGAAAAAC TCAATAAAAG

2300

AGCATGCTCA AGACACTTGA TCGGTACCAG AAATGCAGCT ATGGATCCAT TGAAGTCAAC  
TCGTACGAGT TCTGTGAACT AGCCATGGTC TTTACGTCGA TACCTAGGTA ACTTCAGTTG

2350

AACAAACCTG CCAAAGAACT TGAGGTGTTC TTAATTCAAA TACTATTTTG AGTTCCTATC  
TTGTTTGGAC GGTTCCTTGA ACTCCACAAG AATTAAGTTT ATGATAAAAC TCAAGGATAG

2400

2450

ATATCATTTT AAGAAAGATC TTTTTTTTTT AAAGTTTGTT TTCGTGAAAT ATTTCAGAAC  
TATAGTAAAG TTCTTTCTAG AAAAAAAAT TTTCAAACAA AAGCACTTTA TAAAGTCTTG

2500

AGCTACAGAG AATATCTGAA GCTTAAGGGT AGATATGAGA ACCTTCAACG TCAACAGAGG  
TCGATGTCTC TTATAGACTT CGAATCCCA TCTATACTCT TGGAAGTTGC AGTTGTCTCC

2550

TACATATCTA TCTATACCTC CATATATTTA CTCAATTCTG TATCCATGTA GATTCAATAT  
ATGTATAGAT AGATATGGAG GTATATAAAT GAGTTAAGAC ATAGGTACAT CTAAGTATAA

2600

TGTAGGTGTG TGTGGCTTTT GTTGGTGCAG AAATCTTCTT GGGGAGGATT TAGGACCTTT  
ACATCCACAC ACACCGAAAA CAACCACGTC TTTAGAAGAA CCCCTCCTAA ATCCTGGAAG

2650

GAATTCAAAG GAGTTAGAGC AGCTTGAGCG TCAACTGGAC GGCTCTCTCA AGCAAGTTTCG  
CTTAAGTTTC CTCAATCTCG TCGAACTCGC AGTTGACCTG CCGAGAGAGT TCGTTCAAGC

2700

2750

GTCCATCAAG GTATCTTTAT GCATGGAATC AATGATTCAA ATGAGATTAA TTTGTGTTGT  
CAGGTAGTTC CATAGAAATA CGTACCTTAG TTACTAAGTT TACTCTAATT AAACACAACA

Fig. 1c

2800  
TTAATTATAC TACTATGGTG GTATGATGAT TGTTTGCAGA CACAGTACAT GCTTGACCAG  
AATTAATATG ATGATACCAC CATACTACTA ACAAACGTCT GTGTCATGTA CGAACTGGTC

2850  
CTCTCGGATC TTCAAATAA AGAGCAAATG TTGCTTGAAA CCAATAGAGC TTTGGCAATG  
GAGAGCCTAG AAGTTTATT TCTCGTTTAC AACGAACCTT GGTTATCTCG AAACCGTTAC

2900  
AAGGTATAAT TACAGAATAA ATGCATTTGG TGACTTGCGA TCAATCTCTT TCACAGAGTT  
TTCCATATTA ATGTCTTATT TACGTAAACC ACTGAACGCT AGTTAGAGAA AGTGTCTCAA

2950  
TAAGTTTCTA AATATGTTTT GAAACATCTC TAGTTTTCTT GTTTCTGATT ATAGTCTTTT  
ATTCAAAGAT TTATACAAAA CTTTGTAGAG ATCAAAAGAA CAAAGACTAA TATCAGAAAA

3050  
GGTGAAATGT AAATGTTTAG CTGGATGATA TGATTGGTGT GAGAAGTCAT CATATGGGAG  
CCACTTTACA TTTACAAATC GACCTACTAT ACTAACCACA CTCTTCAGTA GTATACCCTC

3100  
GATGGGAAGG CGGTGAACAG AATGTTACCT ACGCGCATCA TCAAGCTCAG TCTCAGGGAG  
CTACCCCTTC GCCACTTGTC TTACAATGGA TGCGCGTAGT AGTTCGAGTC AGAGTCCCTG

3150  
TATACCAGCC TCTTGAATGC AATCCAACTC TGCAAATGGG GTAAATCTGC CTTGAAAAAT  
ATATGGTCGG AGAACTTACG TTAGGTTGAG ACGTTTACCC CATTTAGACG GAACTTTTTA

3200  
CATCTGCAAA TCAGTTTGTG TACTTAACTA CTAAGATTGT CCTTATTTAA GGTTCTTTAG  
GTAGACGTTT AGTCAAACAC ATGAATTGAT GATTCTAACA GGAATAAAT CCAAGAAATC

3250  
TTGCTTGGTG TAAAGAGGAT CATCAATGTG TGTGAACCTT CTAAGTTGAT GTTTTGGCGA  
AACGAACCAC ATTTCTCCTA GTAGTTACAC ACACTTGGA GATTCAACTA CAAAACCGCT

3350  
TGATGATGAT GATGCAGGTA TGATAATCCA GTATGCTCTG AGCAAATCAC TGCGACAACA  
ACTACTACTA CTACGTCCAT ACTATTAGGT CATACGAGAC TCGTTTAGTG ACGCTGTTGT

3400  
CAAGCTCAGG CGCAGCCGGG AAACGGTTAC ATTCCAGGAT GGATGCTCTG AGAATCATGT  
GTTTCGAGTCC GCGTCGGCCC TTTGCCAATG TAAGGTCCTA CCTACGAGAC TCTTAGTACA

3450  
ACTGTGATGA AGCTCACCCA CAAAAGACCT TATATATATA TAAAGTATAG ATACAAGACT  
TGACACTACT TCGAGTGGGT GTTTTCTGGA ATATATATAT ATTTCAATC TATGTTCTGA

3500  
TGGATTGTA GACATAAGTG GCTAATATAA TGGTCCTGAG GATCTTCTAG ACATTTGTAT  
ACCTAAACAT CTGTATTCAC CGATTATATT ACCAGGACTC CTAGAAGATC TGTAACATA

3550  
CTTTTGGGAA TCCTTGCTTA TATTAAGAAT TCAAATGTGT GGAACCTGTT TTAACACTGA  
GAAAACCCCT AGGAACGAAT ATAATTCTTA AGTTTACACA CCTTGAACAA AATTGTGACT

3650  
ACCATGACAC TGGTTTATTA TCATGTAATG AGAGAAACAT TTGGGTTACA ATGTGATCTC  
TGGTACTGTG ACCAAATAAT AGTACATTAC TCTCTTTGTA AACCCAATGT TACACTAGAG

3700  
TCCTTGACCC AAATACACAA TATAAACCTT ATGCCAAAAT ACAAGCATCA CATATATATA

Fig. 1d



AGGAACTGGG TTTATGTGTT ATATTTGGGA TACGGTTTTA TGTTCTAGT GTATATATAT

3750  
TTCATAAAAG GTTTAAGTAA TCATACAAAT GATGTAAAAA GTTTCATGCC TTGAACAAAA  
AAGTATTTTC CAAATTCATT AGTATGTTTA CTACATTTT CAAAGTACGG AACTTGTTTT

3800  
CACTGCGCCA AAGGCAAATG GTAAGAAACA TGTCAGATTC CTGTGTGCAT CTGTTTTGCT  
GTGACGCGGT TTCCGTTTAC CATTCTTTGT ACAGTCTAAG GACACACGTA GACAAAACGA

3850 3900  
GCTGCTGCTG TTGTTATCTC TCAAGAGGGT TTCCTCAGAA CTCCATAAGC CAAACGTGCA  
CGACGACGAC AACAATAGAG AGTTCTCCCA AAGGAGTCTT GAGGTATTCT GTTGCACGT

3950  
GAGAGACGTT TCCTCATTC CCCATCGTAT ACAATACCAT ATATTGTTAA AAAAAAGATA  
CTCTCTGCAA AGGAGTAAGG GGGTAGCATA TGTTATGTA TATAACAATT TTTTCTCTAT

4000  
TCACAGATCA AATCAATTTG CACATCTCTC TGCTGCCTTG TCAATCTCCT CAGGTCCGGT  
AGTGTCTAGT TTAGTTAAAC GTGTAGAGAG ACGACGGAAC AGTTAGAGGA GTCCAGGCCA

4050  
CAAGGCAGAT CAAGACAGGA TCAATGGCAA CAAGTTACGG TGTTTCGTTG AACTCCATCA  
GTTCCGTCTA GTTCTGTCTT AGTTACCGTT GTTCAATGCC ACAAAGCAAC TTGAGGTAGT

4100  
CCTGCAAATG AGACGAATTC ACAGCAGAGA AAAAAATATT CTTTAGTCAA CATGAATGAG  
GGACGTTTAC TCTGCTTAAG TGTCTCTCTT TTTTTTATAA GAAATCAGTT GTACTTACTC

4150 4200  
AAATAATTCA AATGTTCTGA GTTTCAGGAA GAATGATTAG CCATATTTGT ACTAGACAAG  
TTTATTAAGT TTACAAGACT CAAAGTCCTT CTTACTAATC GGTATAACA TGATCTGTTC

4250  
ACAAGTAAAG ATTTTACGCA TGTGCTTCTA GGGTTGTTGT ACATCTTTCA TTCTATTGAT  
TGTTCAATTC TAAATGCGT ACACGAAGAT CCCAACAACA TGTAGAAAGT AAGATAACTA

4300  
CTCTGGATCA CTCGTCTATT TATGCGTGAT GGTGTCTGAG TCTGACTCTG AAACACTAGT  
GAGACCTAGT GAGCAGATAA ATACGCACTA CCACAGACTC AGACTGAGAC TTGTGTATCA

4350  
AAATGAGAAG CCGAAAAC TGCTTGAAGA ACATGAAAAG TGTTTACCTT TCCACAAACA  
TTTACTCTTC GGCTTTTGAC CGAACCTTCT TGTACTTTTC ACAAATGGAA AGGTGTTTGT

4400  
GGGCAGTTTT CACTTCTCTC CATCCATTCA TAAATGCAAC TAAGGTGGAA ATGGTGAGAA  
CCCGTCAAAA GTGAAGAGAG GTAGGTAAGT ATTTACGTTG ATTCCACCTT TACCACTCTT

4450 4500  
CACTTTGTAA CAATCTTCGG GTTCTCTGAT ATGTATTCTA CAAACACAC GAAATAATCT  
GTGAAACATT GTTAGAAGCC CAAGAGACTA TACATAAGAT GTTTGTGTG CTTTATTAGA

GATACTAAGC TT  
CTATGATTCTG AA

Fig. 1e

-1104  
TGATAGCGCT TCGTTCATCA TGCAGAAGAA ACCAATGTTT CCCCAATCTC  
ACTATCGCGA AGCAAGTAGT ACGTCTTCTT TGGTTACAAA GGGGTTAGAG

-1054  
ACGCGCCTCC TCCTATCTAC CACCACTTGG ACAAATCCCC TTTGCAGTAT  
TGCGCGGAGG AGGATAGATG GTGGTGAACC TGTTTAGGGG AAACGTCATA

-1004  
TCGTTTTTTT TTCCGGACAT TGTACATTCA AAAGCATTCC AAGTGTCTAA  
AGCAAAAAAA AAGGCCTGTA ACATGTAAGT TTTCGTAAGG TTCACAGATT

-954  
TAAACATAAC TAACCACTCC AAGATGCAAA ATCTAGCTAC GACGAACAAA  
ATTTGTATTG ATTGGTGAGG TTCTACGTTT TAGATCGATG CTGCTTGTTT

-904  
TTTTAACTA TAGAGATGAA CTTTAAATTC GGGCATTAAAT TAGTGGAAC  
AAAATTTGAT ATCTCTACTT GAAATTTAAG CCCGTAATTA ATCACCTGA

-854  
TGAGCTATTG ATGATCGAGT TTTCTGACTT TTTGAAGCTT AAGCTTAATT  
ACTCGATAAC TACTAGCTCA AAAGACTGAA AAACCTCGAA TTCGAATTAA

-804  
GAGTTTTATA TACACTATAT AGGCTTGTA TAATATGGAT CAAACAAGAA  
CTCAAAATAT ATGTGATATA TCCGAACATT ATTATACCTA GTTTGTTCTT

-754  
AAATACAAAC TACAAATTGG GAATTGGGTT TTAAACGTT ATCGTTCTAT  
TTTATGTTTG ATGTTTAACC CTTAACCCTA AATTTTGCAA TAGCAAGATA

-704  
TTTAATTCAG GCACGTACCT TTAGAATATC AAGATCCATG TTTCAATATT  
AAATTAAGTC CGTGCATGGA AATCTTATAG TTCTAGGTAC AAAGTTATAA

-654  
TCTGTTGACA AATAAATAAA GATGTCTCAA ATATAAGTTG GGCAACGTAC  
AGACAACTGT TTATTTATTT CTACAGAGTT TATATTCAAC CCGTTGCATG

-604  
GTGTAGACCT AAAAGAGTCG AAACATTGGT ATCTAAGTTA TATATCTACA  
CACATCTGGA TTTTCTCAGC TTTGTAACCA TAGATTCAAT ATATAGATGT

-554  
TGGATTATAT AACAAGACAA CGTTTGTTTT AAAAACTTCA TTGATTTTTC  
ACCTAATATA TTGTCTGTT GCAAACAAAA TTTTGAAGT AACTAAAAAG

-504  
TTAATTAGTA GCAACTAGCA ACTAACTACT CATGGCAAAT AATGGCGTCT  
AATTAATCAT CGTTGATCGT TGATTGATGA GTACCGTTTA TTACCGCAGA

-454  
GCGTGGCAGC CGACTTGGGA GAGAAGGTGT GAGAATGTTT TTAATTTCTG  
CGCACCGTGC GCTGAACCCT CTCTCCACA CTCTTACAAA AATGAAAGAC

-404

Fig. 2a

TGTAAGAGAT GGAAGAGAGA GAAAGAGTAA AGAAGTAGAG AGAGAGATAT  
ACATTTTCTA CCTTCTCTCT CTTTCTCATT TCTTCATCTC TCTCTCTATA

-354

TGTATCACCA AACCTTAATG ATCTCTCACC CTCACAAATT TTCTTATCTT  
ACATAGTGGT TTGGGATTAC TAGAGAGTGG GAGTGTTTAA AAGAATAGAA

-304

TATAGCTTTT ATAGATTCAC AAAAAGTTTT CTTCAGATTC ACAATCTCAT  
ATATCGAAAA TATCTAAGTG TTTTGAAGAA GAAGTCTAAG TGTTAGAGTA

-254

CACAACCCTT CAAAAAGAGA AAAGATCTAA AGAATAAACA AGAGCCCTAA  
GTGTTGGGAA GTTTTCTCT TTTCTAGATT TCTTATTTGT TCTCGGGATT

-204

TATCAAATCA CAACCAAAAA AACCAAAGAA AGCTAATTAA AGTTTTCTCT  
ATAGTTTAGT GTTGGTTTTT TTGGTTTCTT TCGATTAATT TCAAAAGAGA

-154

CTAGCTATTC CTCTCTTTTT CTTGTTCTTG AAAAGTAGGG TTTACTTCAC  
GATCGATAAG GAGAAGAAAA GAACAAGAAC TTTTGATCCC AAATGAAGTG

-104

CAAAAAGATA AGATCTTTCC CCAGAAAAAG CAATACCCAA GTCATGTTTC  
GTTTTTCTAT TCTAGAAAGG GGTCTTTTTC GTTATGGGTT CAGTACAAAG

-54

TGTGTGTCTG TATATAGATA AAACATTACA TACCCTAATA AGGTTACACA  
ACACACAGAC ATATATCTAT TTTGTAATGT ATGGGATTAT TCCAATGTGT

-4

AATAGCTATA AAAGAGGGAA AATAAGATAG GGATTTTTTG GGGTGAGGAA  
TTATCGATAT TTTCTCCCTT TTATCTATC CCTAAAAAAC CCCACTCCTT

47

AGATGGGAAG AGGAAGAGTA GAGCTCAAGA GGATAGAGAA CAAAATCAAC  
TCTACCTTC TCCTCTCAT CTCGAGTTCT CCTATCTCTT GTTTTAGTTG

97

AGACAAGTGA CGTTTGCTAA ACGTAGAAAT GGTTCGTGA AAAAAGCTTA  
TCTGTTCACT GCAAACGATT TGCATCTTTA CCAAAGCACT TTTTTCGAAT

147

TGAGCTTTCT GTTCTCTGCG ATGCTGAAGT CTCTCTCATC GTCTTCTCCA  
ACTCGAAAGA CAAGAGACGC TACGACTTCA GAGAGAGTAG CAGAAGAGGT

197

ACCGTGGCAA GCTCTACGAG TTCTGCAGCA CCTCCAAGTA CTTCTCTTTC  
TGGCACCGTT CGAGATGCTC AAGACGTCGT GGAGGTTTCAT GAAGAGAAAG

247

TTTATACACT TATTAGATCT GTGTGTAGAT CTTTCATTTT TTCTAGTCTT  
AAATATGTGA ATAATCTAGA CACACATCTA GAAAGTAAAA AAGATCAGAA

297

GTGATGAGTT TTATCTTTCT TGATTGCTTT TTAACAAAAT ACTTGATATA

Fig. 2b

09869582, 022802

CACTACTCAA AATAGAAAGA ACTAACGAAA AATTGTTTGA TGA ACTATAT  
347  
TTTTCAGTTT CTTAATCTGA CTCTAATTAG GTTTTGATTA ATAGGAAGGA  
AAAAGTCAAA GAATTAGACT GAGATTAATC CAAAATAAT TATCCTTCCT  
397  
AATAAATCCA GGTACCTTTC AAGGTGAATT G-----GAG ATCTGATCTT  
TTATTTAGGT CCATGGAAAG TTCCACTTAA C-----CTC TAGACTAGAA  
447  
AATTTAATCA TCATGTCAAA TTCTTAGGGA TTTAATTGCA ATCTATTTTT  
TTAAATTAGT AGTACAGTTT AAGAATCCCT AAATTAACGT TAGATAAAAA  
497  
AGATTTATCG GAGCTAGGAA AGTATCATAA TGATATACTA TTATTATCAT  
TCTAAATAGC CTCGATCCTT TCATAGTATT ACTATATGAT AATAATAGTA  
547  
GTAATTTTCA TGTCTCTACA CGGATATATA TGTGATTAGA ACTTGGTAAA  
CATTAAAGTA ACAGAGATGT GCCTATATAT AACTAATCT TGAACCATTT  
597  
GTAAACTAAA GATTCACAGT CTTCAATGAA ATTGAAAAGA TCCAACGTAG  
CATTTGATTT CTAAGTGTC AAGTTACTT TAACTTTTCT AGGTTGCATC  
647  
AATAATTAGT GGTTCATGC ATTAACCAGT CTAATTAAAG CTCATGCAGA  
TTATTAATCA CCAAGGTACG TAATTGGTCA GATTAATTTT GAGTACGTCT  
697  
CATTTAAGCA CCACATGAAT TTAATATCTT TTTAATTAAG GGATCTTCTT  
GTAAATTCGT GGTGTACTTA AATTATAGAA AAATTAATTC CCTAGAAGAA  
747  
TTTATAAATT TTCTTTTGT AGCTTTTAAA ATTTTAGTTT GTTCATTAAA  
AAATATTTAA AAGAAAACAA TCGAAAATTT TAAAATCAAA CAAGTAATTT  
797  
ATTTATAGAT CCTCCTCTCC TGATTGTGT TTTCCGATCC TTTCCAGCAT  
TAAATATCTA GGAGGAGAGG ACTAAACACA AAAGGCTAGG AAAGGTCGTA  
847  
GCTCAAGACA CTGGAAAGGT ATCAGAAGTG TAGCTATGGC TCCATTGAAG  
CGAGTTCTGT GACCTTTCCA TAGTCTTCAC ATCGATACCG AGGTAACCTC  
897  
TCAACAACAA ACCTGCTAAA CAGCTTGAGG TTTAATCTCC AACATCTCTT  
AGTTGTTGTT TGGACGATTT GTCGAACCTC AAATTAGAGG TTGTAGAGAA  
947  
CGATCTTAAT TATTTATCCT TTTTAAATT TATCTAAAGA AAATGTTTGA  
GCTAGAATTA ATAAATAGGA AAAAATTAAA ATAGATTCTT TTTACAACT  
997  
TTTTGAGACA AAAGCCCTTC AAAGTTTCTT ACATAGATAT TCAATTGTCT  
AAAACTCTGT TTTCGGGAAG TTTCAAAGAA TGTATCTATA AGTTAACAGA

Fig. 2c.

1047  
ATTATCTTCG CAATTTTCAG AACAGCTACA GAGAGTACTT GAAGCTGAAA  
TAATAGAAGC GTTAAAAGTC TTGTCGATGT CTCTCATGAA CTTTCGACTTT

1097  
GGTAGATATG AAAATCTGCA ACGTCAGCAG AGGTATATAC ATTAATGTGG  
CCATCTATAC TTTTAGACGT TGCAGTCGTC TCCATATATG TAATTACACC

1147  
ATGATGATCA TTTATAAACA GCATATATAT ATATATATAT ATATATATAT  
TACTACTAGT AAATATTTGT CGTATATATA TATATATATA TATATATATA

1197  
ATATAGAAAG TATTGATCAT GAAAGTGTGT TGCAGCAGAA ATCTTCTTGG  
TATATCTTTC ATAAC TAGTA CTTTCACACA ACGTCGTCTT TAGAAGAACC

1247  
AGAGGATCTT GGACCTCTGA ATTCAAAGGA GCTAGAGCAG CTTGAGCGTC  
TCTCCTAGAA CCTGGAGACT TAAGTTTCCT CGATCTCGTC GAACCTCGCAG

1297  
AACTAGACGG CTCTCTGAAG CAAGTTCGCT GCATCAAGGT GATTTACTTC  
TTGATCTGCC GAGAGACTTC GTTCAAGCGA CGTAGTTCCA CTAAATGAAG

1347  
TGTACATACA CTGAAAGATT CACACAAATC TTTCTCTATA TATAGACTGA  
ACATGTATGT GACTTTCTAA GTGTGTTTAG AAAGAGATAT ATATCTGACT

1397  
GACACATGCA TGAAATGTTT TTGATGCGTG AGGTTATCTG AAAATGCCTC  
CTGTGTACGT ACTTTACAAA AACTACGCAC TCCAATAGAC TTTTACGGAG

1447  
TTCTTTTTTG CAGACACAGT ATATGCTTGA CCAGCTCTCT GATCTTCAAG  
AAGAAAAAAC GTCTGTGTCA TATACGAACT GGTGAGAGA CTAGAAGTTC

1497  
GTAAGGAGCA TATCTTGCTT GATGCCAACA GAGCTTTGTC AATGAAGGTA  
CATTCCTCGT ATAGAACGAA CTACGGTTGT CTCGAAACAG TTACTTCCAT

1547  
TATGATGATG TTTCTCTCTC TCTCCTCCAG TTTCTATTTA TAGATGGAAA  
ATACTACTAC AAAGAGAGAG AGAGGAGGTC AAAGATAAAT ATCTACCTTT

1597  
CTTTAAATAG TCCAATTTAT ATATATGAGT CTAAATTTCA CATTCTTCAA  
GAAATTTATC AGGTTAAATA TATATACTCA GATTTAAAGT GTAAGAAGTT

1647  
CTGCTACATG TTTCTTTTGT ATTATTTCTA TGATATCTTC AGGAAAGTTT  
GACGATGTAC AAAGAAAACA TAATAAAGAT ACTATAGAAG TCCTTTCAAA

1697  
GAAAAATATT GTGTTTTGTT TAGCTGGAAG ATATGATCGG CGTGAGACAT  
CTTTTATATA CACAAAACAA ATCGACCTTC TATACTAGCC GCACTCTGTA

Fig. 2d

1747  
CACCATATAG GAGGAGGATG GGAAGGTGGT GATCAACAGA ATATTGCCTA  
GTGGTATATC CTCCTCCTAC CCTTCCACCA CTAGTTGTCT TATAACGGAT

1797  
TGGACATCCT CAGGCTCATT CTCAGGGACT ATACCAATCT CTTGAATGTG  
ACCTGTAGGA GTCCGAGTAA GAGTCCCTGA TATGGTTAGA GAACTTACAC

1847  
ATCCCACTTT GCAAATTGGG TAAATCAAAC AACTTTTCTT GCTTTAAGAC  
TAGGGTGAAA CGTTTAACCC ATTTAGTTTG TTGAAAAGAA CGAAATTCTG

1897  
ATCAACTTAG GTTATAAACA GTTAGCAGTT TGCTTTAAGC CCAACATTGT  
TAGTTGAATC CAATATTTGT CAATCGTCAA ACGAAATTCG GGTGTAAACA

1947  
CTTTGTTTCA TAGAGGCTTT GGTAAAACT CGTGTGTTT AGTCTAAGGA  
GAAACAAAGT ATCTCCGAAA CCAATTTTGA GCACAACAAA TCAGATTCTT

1997  
TTCAGCACTT TGATGTCTGA AGTATGGAAT ATCAATCTCT CAGACTTGAA  
AAGTCGTGAA ACTACAGACT TCATACCTTT TAGTTAGAGA GTCTGAACCT

2047  
AATGTGGGTT TCTATTGTTG ACTTCGAAAC TATGTTGTTG TGGTGTGCA  
TTACACCCAA AGATAACAAC TGAAGCTTTG ATACAACAAC ACCACAACGT

2097  
AACAGATATA GCCATCCAGT GTGCTCAGAG CAAATGGCTG TGACGGTGCA  
TTGTCTATAT CGGTAGGTCA CACGAGTCTC GTTTACCGAC ACTGCCACGT

2147  
AGGTCAGTCC CAACAAGGAA ACGGCTACAT CCCTGGCTGG ATGCTGTGAG  
TCCAGTCAGG GTTGTTCCTT TGCCGATGTA GGGACCGACC TACGACACTC

2197  
CGATACTTCT TCCCCAATA AAGATCTTAA GCAAGTACTG GTGGGGTCTT  
GCTATGAAGA AGGGGGTTAT TTCTAGAATT CGTTCATGAC CACCCAGAA

2247  
CGTGGTGTGA TCTTAGATCT TATGCATATG AATAATAATG TTATTGCACA  
GCACCACACT AGAATCTAGA ATACGTATAC TTATTATTAC AATAACGTGT

2297  
AGACTTTTGC TTTTGTAGAC ACAAGTGGCT ATAGCTGTAA TAGCCTTCAA  
TCTGAAAACG AAAACATCTG TGTTCAACCGA TATCGACATT ATCGGAAGTT

2347  
CATCTCTCTT CTGTTTCAGG ATTTGTTTGT GCCTATTGTA ATTGCTTATA  
GTAGAGAGAA GACAAAGTCC TAAACAAACA CGGATAACAT TAACGAATAT

2397  
TATGTATGGT TTGTATAATG TGTGAAATGT TAACATCGAC CATGTCTCAT  
ATACATACCA AACATATTAC AACTTTTACA ATTGTAGCTG GTACAGAGTA  
CTGGTGAAGA TCTTATCCTG TCTATGCATG ATACCAAAA

Fig. 2e

GACCACTTCT AGAATAGGAC AGATACGTAC TATGGTTTT

09869582 0023578

Fig. 2f

Sequence Range: 1 to 14940

50  
TAAATCTGG AAGTTTCCAG CCCTGATAAT GTTGCAGAAT AAATTAGTGC GCAGTAAGTC  
ATTTTAGACC TTCAAAGGTC GGGACTATTA CAACGTCTTA TTTAATCACG CGTCATTACG

100  
TCCAAAAAGA GAGAACTAC AAATAAATAA ACCAAGTCAA ATTCATTAAAC AAGGAGAACA  
AGGTTTTTCT CTCTTTGATG TTTATTTATT TGGTTCAGTT TAAGTAATTG TTCCTCTGT

150  
GCATGAAATG TTTCCCAAAC ACACAAAATC TTGACTAGCC AACAGCGCTT CAAATGAGGA  
CGTACTTTAC AAAGGGTTTG TGTGTTTAG AACTGATCGG TTGTCGCGAA GTTTACTCCT

200  
AGTAACTAAT TTCAGTAGCT TGGGTATGGT GAAGTATAAT TACCTTCCAC CACACATATC  
TCATTGATTA AAGTCATCGA ACCCATACCA CTCATATTA ATGGAAGGTG GTGTGTATAG

250 300  
CGTAGCCTAT CACCCCAACG ATAATGATCA AACCATAGTT TCTACCACCT GTACATTGAA  
GCATCGGATA GTGGGGTTGC TATTACTAGT TTGGTATCAA AGATGGTGGA CATGTAACCT

350  
GGAAAGTGTT AACTGTTTTC TTCCGAATTT AGATCAACAG TAAACAAAGA ATGGTGTAC  
CCTTTCACAA TTGACAAAAG AAGGCTTAA TCTAGTTGTC ATTTGTTTCT TACCACAATG

400  
TCTAAGTCTC TAATGTAATG CCTTCCTAAA TGCTACAAAG AAAAGCCACT TATCAGAACA  
AGATTCAGAG ATTACATTAC GGAAGGATTT ACGATGTTTC TTTTCGGTGA ATAGTCTTGT

450  
AAGTATGTCT TGTTTGATGC GAGAAAAGTA GCAAAAGAGA ATAAAACCTG AAATATAATT  
TTCATACAGA ACAAACCTACG CTCTTTTCAT CGTTTTCTCT TATTTTGGAC TTTATATTAA

500  
TCAAAATACA ATGTCTAGAA ATCTAAGTGT GCAAAATCCTT TATTCAAGTT TCATATCAAA  
AGTTTTATGT TACAGATCTT TAGATTCACA CGTTTAGGAA ATAAGTTCAA AGTATAGTTT

550 600  
CCAATTTTGA CATTTCTAGT GCAGAACAGA AAACAAAAC TCAATATAAA AAAATATAAA  
GGTTAAACT GTAAAGATCA CGTCTGTCT TTTGTTTTGA AGTTATATTT TTTTATATTT

650  
AACTCCAGAG GACCTGATCC TGAAGGTGAA ACAATGGTGA TAGGTCTGTT TGACCCAGC  
TTGAGGTCTC CTGGAAGTAG ACTTCCACTT GTTACCCT ATCCAGACAA ACTGGGGTCG

700  
AACTGTATCT CATGCCTAAG ACTGTAAACC TACAAAATA AATAGAGCTC AGGCAAGAAA  
TTGACATAGA GTACGGATTC TGACAATTGG ATGTTTTTAT TTATCTCGAG TCCGTCTTTT

750  
CTATTGATTC ACGATAAATC TATGTCCTCA GCAAGTCTAT ATTATCCAGC TCCATCCGAT  
GATAACTAAG TGCTATTTAG ATACAGGAGT CGTTCAGATA TAATAGGTG AGGTAGGCTA

800  
AGCTTATCAT CGCCAATAGA TTAATGTGAA ACTTACCTGG GCCACAAGTA CATCATCGTG  
TCGAATAGTA GCGGTATCT AATTACACTT TGAATGGACC CGGTGTTCAT GTAGTAGCAC

850 900  
GGGTTTGCTA GCTGATTGTC TAGGTTGCTC TTGTTTCAGT TGCCTGAATA CCATCTGTCC  
CCCAAACGAT CGACTAAACG ATCCAAGCAG AACAAAGTCA ACGGACTTAT GGTAGACAGG

Fig. 3a



950  
ACATAAACAA AACCCATTGC CTCATTTTGC CAAACCGCAT CATAACATG TGAAGTCGCC  
TGTATTTGTT TTGGGTAACG GAGTAAACG GTTTGGCGTA GTATGTGTAC ACTTCAGCGG

1000  
AAAGCTTTTG CACAATATAG AAATTAGAAT ACCTTAAAAG CACCAGAAAC CAAATTGGAG  
TTTCGAAAAC GTGTTATATC TTTAATCTTA TGGAAATTTT GTGGTCTTTG GTTTAACCTC

1050  
ACATCTGGTA AGCCCCCTTC TTTAGAAAAT GCTGATCCAA TAAGACCTTA AAGTAACATT  
TGTAGACCAT TCGGGGGAAG AAATCTTTTA CGACTAGGTT ATTCTGGAAT TTCATTGTAA

1100  
TGCAAAAATC ACAGTATAGT TAGTAATTGC AGTAACTTGG ACGAACATTA AGCATGTACA  
ACGTTTTTAG TGTCAATATCA ATCATTAAACG TCATTGAACC TGCTTGTAAT TCGTACATGT

1150 1200  
CGAAATCAAT CGACTCAGCA AGTTCACAAT AATTGTACTA GTAGGTGCAT TCACAGAGAA  
GCTTTAGTTA GCTGAGTCGT TCAAGTGTTA TTAACATGAT CATCCACGTA AGTGTCTCTT

1250  
ACTAAACATA AACTTCTCCT CAGATGTATT CAGAGAATAG CTATACTCCA ATAAAGTCTT  
TGATTTGTAT TTGAAGAGGA GTCTACATAA GTCTCTTATC GATATGAGGT TATTTAGAA

1300  
AAACTTTGAG CCAGTCAAGT ACACTGATCA AAGGGTTTAT GAAAAACACT AACTTCTTAT  
TTTGAAACTC GGTCAAGTTCA TGTGACTAGT TTCCCAAATA CTTTTTGTGA TTGAAGAATA

1350  
CCTCTAATTG CGATTACCCA TAGACGAAAC CAATAAAAAA GCAATGGAGA ACTAGAGCAC  
GGAGATTAAC GCTAATGGGT ATCTGCTTTG GTTATTTTTT CGTTACCTCT TGATCTCGTG

1400  
AGTCACTACA AGAAATACCC TATAAAAGTA CCGACCTGCA CCGATGAGGA TGGTGAGCTT  
TCAGTGATGT TCTTTATGGG ATATTTTCAT GGCTGGACGT GGCTACTCCT ACCACTCGAA

1450 1500  
CCCAGCGGA AGAGCCATGG CTAGAGACGA GCTTATACGG CGAAGAACTA AGATGGCAAA  
GGGCTCGCCT TCTCGGTACC GATCTCTGCT CGAATATGCC GCTTCTTGAT TCTACCGTTT

1550  
CGAATCCGCG TGAGAATATC TAAGAGAGTA TTGGTAAGAG AGAGCTGCAG GAACGTACCG  
GCTTAGGCGC ACTCTTATAG ATTCTCTCAT AACCATTCTC TCTCGACGTC CTTGCATGGC

1600  
GTGAAACAGA GCGTTTTTTT GGGACGATGA AGTGAGGCAG CGAGAGAGAT ACGACGTGCG  
CACTTTGTCT CCGCAAAAAA CCCTGCTACT TCACTCCGTC GCTCTCTCTA TGCTGCACGC

1650  
ACTATATTGT TCGCTTGTG AGGCAACAAA ACAGAGTTGC TTCTAAACC CGAACCGAAA  
TGATATAACA AGCGAACAAAC TCCGTTGTTT TGTCTCAACG AAGATTTTGG GCTTGGCTTT

1700  
TGTCGGGTCT GATTCCGTCT AAATCACGAT TAGGTTCTGT TTAACCTA GGAGGCAATA  
ACAGGCCAGA CTAAGCCAGA TTTAGTGCTA ATCCAAGCAA AATTTTGGAT CCTCCGTTAT

1750 1800  
ACCGGACGGA TCATAAATTC ATAATAGAGA CAGACAAATT GGTCCATTAT TAAATCACT  
TGGCTGCCT AGTATTTAAG TATTATCTCT GTCTGTTTAA CCAGGTAATA ATTTTAGTGA

1850  
TGGGCATTTG GGGATGATTC AAATGCCCAA GTTTTCTCAA ATTTGGACGA TTCATTACC

Fig. 3b

ACCCGTAAAC CCCTACTAAG TTTACGGGTT CAAAAGAGTT TAAACCTGCT AAGTAAGTGG

1900  
TAAGACATAC TTGAGCAACA ACAAAGTGAA GTCCACTGTC ATATCTTATG TCTCAAAAAG  
ATTCTGTATG AACTCGTTGT TGTTCACCT CAGGTGACAG TATAGAATAC AGAGTTTTC

1950  
TATTGAAATG TGTCAATTGA TATTGGAGAG GCACACTAGC TAAGGGATTA TTCAATCAAT  
ATAACTTTAC ACAGTTAACT ATAACCTCTC CGTGTGATCG ATTCCCTAAT AAGTTAGTTA

2000  
TTCCAGCAAT TTAATTAAAC TTATTTGTAG TGAAAGTGGG AAGATAAAG ATCTCACCC  
AAGGTCGTTA AATTAAATTG AATAAACATC ACTTTCACCC TTCTATTTTC TAGAGTGGGA

2050  
CACATGTTCA AAAAAAAG TTGAAATGG AAGTAATTCA ACATGTAGCA TAGAGCCCA  
GTGTACAAGT TTTTTCCTTC AACTTTTACC TTCATTAAGT TGTACATCGT ATCTCGGTT

2100  
2150  
ATATGTCTCA TTTTTCCTAT CCATATAATC TCAAATCCTC TTAATCTACT CTAAACATAT  
TATACAGAGT AAAAAAATTA GGTATATTAG AGTTTAGGAG AATGAATGAA GATTGTGATA

2200  
GGTTCCCAT ATCATAACAA TGCTATGTGA ACATGGCCGG TTCTAAAGGA AGCCAAGTGC  
CCAAGGGTAT TAGTATTGTT ACGATACAAAT TGTACCGGCC AAGATTTTCT TCGGTTACAG

2250  
AGCAACTGCC TTACGCCTCT ACGTGTAAAA ATGAAATGA AGACCACTGA CCACTTCTAT  
TCGTTGACGG AATGCGGAGA TGCACAATT TACTTTTACT TCTGGTGACT GGTGAAGATA

2300  
TAAAGCTTCA TTCCTAGTGA TATAATTACA CATTTCCTTA AGGATTTATG AGTAGTGATT  
ATTTGCAAGT AAGTGATCAC ATATTAATGT GTAAAAAAT TCCTAAATAC TCATCACTAA

2350  
GAGGCCCAT TGTTCGTATG TTTGTTTTC TTAATATATC ATTACTTGAC TATAAGAGTT  
CTCCGGGTAT ACAAACATAC AAACAAAAAG AATGATATAG TAATGAAGT ATATTCTCAA

2400  
2450  
GGTTTCCTAT TCCATCTCT TTTCTAACAG CCTATATATG TAAAAATCTA AGCAAAATTT  
CCAAAGGATA AGGTAAAGAGA AAAGATTGTC GGATATATAC ATTTTATAGT TCGTTTAA

2500  
CTTGTCAAGA GGATGATTGT ACATTTGTAC TTGGTTATCT CGCCCCGCC CAAAACATAC  
GAACAGTTCT CCTACTAACA GTAAACATG AACCAATAGA GCGGGGCCG GTTTTGTATG

2550  
CTAAGGCCAG GTGCTATATC CTCAACCTGC TTTGGCATT CACAATCTAC GAACTTTGGC  
GATTCCGTC CACGATATAG GAGTTGGACG AAACCGTAAG TAGTTAGATG CTTGAAACCG

2600  
GTGAAACGGT GACAAGATTA ACAAGATTCA CTCTCAACTA CGATGTTCTA CTATCTCAA  
CACTTTGCCA CTGTTCTAAT GTTCTAAGT GAGAGTTGAT GCTACAAGAT GATAGAGTTT

2650  
TCTTTAAAA AGTGGATCAA ACTGTCAAAA GTCTAGTTCG ATGGACTAGC TTCAACACTC  
AGAAATTTT TCACCTAGTT TGACAGTTT CAGATCAAGC TACCTGATCG AAGTTGTGAG

2700  
2750  
CTCAAATCT AGTTCGATGG ACTATATATT CTCTTCTGAT GCTATCCTTA TCTTGGATTA  
GAGGTTTGA TCAAGCTACC TGATATATAA GAGAAGACTA CGATAGGAAT AGAACCTAAT

Fig 3c

2800  
GGCATCTAAA CTATGGTTTT AATGGTGTCA TGAGGTTTTA CAACTTACAA GGATGAAAGT  
CCGTAGATTT GATACCAAAA TTACCACAGT ACTCCAAAAT GTTGAATGTT CCTACTTTCA

2850  
TATTTACTCC CAGTCACTAT CTTAATCAAA TGACAAAATG TTAAGTAGTT TGAGTGCTTA  
ATAAATGAGG GTCAGTGATA GAATTAGTTT ACTGTTTTAC AATTGATCAA ACTCACGAAT

2900  
TATATTAGTT ATGAATCTGA AATTTATTAG TGTGTACATA AGTGATACAA CACTTAAATA  
ATATAATCAA TACTTAGACT TTAAATAATC ACACATGTAT TCACTATGTT GTGAATTTAT

2950 3000  
ACATCTACAT GAGTTTTTAA ATAACATAAT AATCCATTAT AGTAGTTTAC GGCATAAGGT  
TGTAATGTA CTCAAAATT TATTGTATTA TTAGGTAATA TCATCAAATG CCGTATTCCA

3050  
ATGAACCAAA TTTTTCATTG CACGCTGAAA AGTGAAAACC TTTAAAATGC ATAATGACTA  
TACTTGTTT AAAAAGTAAC GTGCGACTTT TCACTTTTGG AAATTTTACG TATTACTGAT

3100  
AGAGTCTATG ACAACAGTAA CTTACTATAT ATTAGAGGAG GGGTGAAAAA AAAAGTAGAG  
TCTCAGATAC TGTGTGCATT GAATGATATA TAATCTCCTC CCCACTTTTT TTTTCATCTC

3150  
AGACTGGTCC AAAAATTAA CCCCACTCAA TAAACCCAGA CGTGAAGTGT TTGACGATAA  
TCTGACCAGG TTTTGAATT GGGGTGAGTT ATTTGGGTCT GCACTGAACA AACTGCTATT

3200  
CTCCATCTTT CTATTTTGGG TAACGAGGTC CCCTTCCCAT TACGTCTTGA CGTGGACCTT  
GAGGTAGAAA GATAAAACCC ATTGCTCCAG GGAAGGGTA ATGCAGAACT GCACCTGGGA

3250 3300  
GTCCGTCTAT TTTTAGCAGA TTAATCCAAC GGTTCCTTAT CTTTCTTCGA CCCTTCACGA  
CAGGCAGATA AAAATCGTCT AATTAGGTTG CCAAGAATAA GAAAGAAGCT GGAAGTGCT

3350  
CATTGCCCTA AAGCCGTCCG ATTCTCATCT CACGCCCAAT GGACCACATA TATCACCAGT  
GTAACGGAGT TTCGGCAGGC TAAGAGTAGA GTGCGGGTTA CCTGGTGTAT ATAGTGGTCA

3400  
ACTCCGCAAC TTAGCTGTG TGTAGGATTT CACGTGGCAT TTATTTGTTT TAGTTGTAG  
TGAGGCGTTG AATCGACAGC ACATCCTAAA GTGCACCGTA AATAACAAG ATCAAACATC

3450  
TGCAACATT GCAAGTTGAT ATGGTCCCCT ATCGATCACC GTCGTCTCTT TAGCTTCACA  
ACGTTTGTA CTTTCAACTA TACCAGGGGA TAGCTAGTGG CAGCAGAGAA ATCGAAGTGT

3500  
TCGAGATTCT TCTTCTTTT CTACGTGTAA TAGCATTTT GATTTTGAGA ATTTCTTTAG  
AGCTCTAAGA AGAAGAAAG GATGCACATT ATCGTAAAA CTAAACTCT TAAAGAAATC

3550 3600  
AACCCTTGA TCTCTCATCG TTGGTTGATC CATCCATCCA AATGGGACCT GTGTGTGCTC  
TTGGCAACCT AGAGAGTAGC AACCAACTAG GTAGGTAGGT TTACCCTGGA CACACACGAG

3650  
CATCCAGGGC ATATGATCCC AAAGCCAAAA GAGTATTTCC AAGTGCTTTC TTTCTTTCTT  
GTAGGTCCCG TATACTAGGG TTTCCGTTTT CTCATAAAGG TTCACGAAAG AAAGAAAGAA

3700  
TCTTTCTTTC TTACTAACCT TTTTTTTTCT TATGCTTTAG ACTAAGAAAT TTATTCGGCC

Fig. 3d

AGAAAGAAAG AATGATTGGA AAAAAAAGA ATACGAAATC TGATTCTTTA AATAAGCCGG

3750  
ATATCCACTT TTACGAATAT ACTTCTTACA AGATCTAGAT TTTTTTGAGT TAATTCGGTG  
TATAGGTGAA AATGCTTATA TGAAGAATGT TCTAGATCTA AAAAACTCA ATTAAGCCAC

3800  
TATATAACAT TGGCATGGAC TGCAATTAAG TAATGGTAAT GTGATCATGA TGCGATGTGT  
ATATATTGTA ACCGTACCTG ACGTTAATTC ATTACCATTA CACTAGTACT ACGCTACACA

3850 3900  
CGTTATCAGT AGTATAATAT TGATGGGCTA CCCTGGAAAA CAAATTACG TGTTATATGT  
GCAATAGTCA TCATATTATA ACTACCCGAT GGGACCTTTT GTTTTAATGC ACAATATACA

3950  
ACACAATTTG GTAGAACCGT AGAAATTAAA CTGAATAAAA CTTTCTATAA TGTTCAAAAT  
TGTGTTAAAC CATCTTGGCA TCTTTAATTT GACTTATTTT GGAAGATATT ACAAGTTTTA

4000  
TATATGGTAC AGATTAATAC GGAAAAACAT TCACGCTTTA CGTAACAATT AAGTGGAAG  
ATATACCATG TCTAATTATG CCTTTTGTGA AGTGCGAAAT GCATTGTAA TTCACCTTTT

4050  
TAAAATTATC CCAAAAATAT TTATATCACA TCATTGTTAT ATTTCTAAGT TTTTTATAT  
ATTTTAATAG GGTTTTTATA AATATAGTGT AGTAACAATA TAAAGATTCA AAAAAATATA

4100  
CTCTAATGGT ATATGTTTTA CAGATTGTTT TTTGGGAAAA TTCTTAAAGA GACTTGAAGA  
GAGATTACCA TATACAAAAT GTCTAACAA AAACCCTTTT AAGAATTCT CTGAACCTCT

4150 4200  
ATGTTTTTTT TTTATTTTCT TGAAATGTTT GACACTTGAA ACCGTTTAAA AACTCAAATA  
TACAAAAAAA AAATAAAGA ACTTTACAA CTGTGAACCT TGGCAAATTT TTGAGTTTAT

4250  
TAGTATATAT CATTGTTGGT CTCATACCTT GTAATTCACC ACATATATTA TCAATGGGGA  
ATCATATATA GTAACAACCA GAGTATGGAA CATTAAAGTG TGTATATAAT AGTTACCCCT

4300  
AGATTTGAAA ATTTTGGGG GATCACAAAA CGAAGGAAAG AGTACAAAA GAGAAGGAAA  
TCTAACTTT TAAAAACCC CTAGTGTTTT GCTTCCTTC TCATGTTTT CTCTTCCTT

4350  
AGATAGAAGA TATATGTTTT TAACTTCATT GGTATGACAT CAATAAATA ATAGTTGAAT  
TCTATCTTCT ATATACAAA ATTGAAGTAA CCATACTGTA GTTATTTATT TATCAACTTA

4400  
GTACTTTAGT TTCTCTTTTG GTTTAATGCA CATCATCTCG ATCAATTGTC ATCATCTTAC  
CATGAAATCA AAGAGAAAAC CAAATTACGT GTAGTAGAGC TAGTTAACAG TAGTAGAATG

4450 4500  
ATTGAATTAT ACGACCAGAT CTGATAACAA GTGAATTCGT ACTTGCCCTT CCCTTCTTC  
TAACTTAATA TGCTGGTCTA GACTATGTGT CACTTAAGCA TGAACGGGAA GGGAAAGAAG

4550  
TCATACGTCC TTCTAACTAA TTTTGATTGT AACTTATAAT TATATAACCA TATTTAATTT  
AGTAGCAGG AAGATTGATT AAACTAACA TTGAATATTA ATATATTGGT ATAAATTAAA

4600  
TATTTTATCT AAAACCAATT GAAGCAAATT AAAATATCAT AAATCTTGAG TCCACATGA  
ATAAAATAGA TTTTGGTTAA CTTCGTTTAA TTTTATAGTA TTTAGAACTC AGGTGTACT

Fig. 3e

4650  
AGACAATATA TAAAACTCGT GCAAATTGCT TTAAAATGCT TCTATGAGAC CATGACCAAG  
TCTGTTATAT ATTTTGAGCA CGTTTAAACG AATTTTACGA AGATACTCTG GTACTGGTTC

4700  
TGAGATTAAT AAGCGATTCA ATGTGCAAAT CAAAAGAGAA AAGAAGCTAA TGGGTTTAAA  
ACTCTAATTA TTCGCTAAGT TACACGTTTA GTTTTCTCTT TTCTTCGATT ACCCAAATTT

4750 4800  
TATAACCAAA CAGAATAATA ATGCTATGTT TAGTTTTTCT AATTGAATCA TACCTTTGTG  
ATATTGGTTT GTCTTATTAT TACGATACAA ATCAAAAAGA TTAACCTAGT ATGGAAACAC

4850  
TCCATCACCT ACTTACCGGT CAGAATAAAG CAATTACGTC TGCAACCAAA AAGCACTAAG  
AGGTAGTGGA TGAATGGCCA GTCTTATTTC GTTAATGCAG ACGTTGGTTT TTCGTGATTC

4900  
ACTTTCGGTC AGACATGATC TCTAACATCG GACGAACCCCT AAGATAACCA AAATAAACTA  
TGAAAGCCAG TCTGTACTAG AGATTGTAGC CTGCTTGGGA TTCTATTGGT TTTATTTGAT

4950  
TATCTTATAT TCAAATCTCT GTTTATTTTA TCCATTTATG TTTTCTTTCT TTCCATAAT  
ATAGAATATA AGTTTAGAGA CAAATAAAAT AGGTAAATAC AAAAGAAAGA AAGGGTATTA

5000  
TTTTTTTGTG TCTCATCAGA CTCTCTTACC AAACCTGAATT TATCAACATG GTTTTTTTTT  
AAAAAACAC AGAGTAGTCT GAGAGAATGG TTTGACTTAA ATAGTTGTAC CAAAAAATA

5050 5100  
TGGCCACATC AAAATGGTGG TTTATAAAGT AGACTAATAC AAAAGACATT TCTGTTAATT  
ACCGGTGTAG TTTTACCACC AAATATTTCA TCTGATTATG TTTTCTGTAA AGACAATTA

5150  
TCACTAACAA AAATAATCTT AGCAGTACTA TAGATTGGAA AAGGAAAAGC AAATCTAGCA  
AGTGATTGTT TTTATTAGAA TCGTCATGAT ATCTAACCTT TTCCTTTTCG TTTAGATCGT

5200  
GTAAGATTTA TCAAACTAG CAGTAAGAGT TTTAGATATC ATGAAAACAT CACAAACGAG  
CATTCTAAAT AGTTTGTATC GTCATTCTCA AAATCTATAG TACTTTTGTA GTGTTTGCTC

5250  
TAGTGTTTTA CTTTACATTT TTAACCAATC ACAAGGGTAG TTCCGTAAGT TGGGAAAATC  
ATCACAAAAT GAAATGTAAA AATTGGTTAG TGTTCCTATC AAGGCATTCA ACCCTTTTAG

5300  
GTACGAGGCT TCACCTAGTT AAGGTTAGGT CACATGATTC CCTGAACCTG ATTTTATAAG  
CATGCTCCGA AGTGATCAA TTCCAATCCA GTGTACTAAG GGAATTGAGC TAAAATATTC

5350 5400  
TAAAAAAGAA AAATTTATAA AATCAAAATT TTTTATATAA AAAAATCAGG TGGATTTATC  
ATTTTTTCTT TTAAATATT TTAGTTTAA AAAATATATT TTTTATAGTC ACCTAAATAG

5450  
AGACCCTACC ATCGAGATGT CGACACGTGT CCAAACCTCAT TCATTGCCCT ACTATTTTCT  
TCTGGGATGG TAGCTCTACA GCTGTGCACA GGTTTGAGTA AGTAACGGGA TGATAAAGA

5500  
GTTTAGGGTT GCAATCACTC ATCGCACACG CGCCATCTCC ACCTTCCATT ATTAATCTCT  
CAAATCCCA CGTTAGTGAG TAGCGTGTGC GCGGTAGAGG TGAAGGTAA TAATTAGAGA

5550  
CATTTTCAAC ATCACACTCT TACGAATCAT ACGATTTTAA TATCTCTGTC TCTCTCAACG

Fig. 3f

09869582-023578

GTAAAAGTTG TAGTGTGAGA ATGCTTAGTA TGCTAAAATT ATAGAGACAG AGAGAGTTGC

5600

TATTAAATAA AAATGGTTT AAATGTTAGG GTTTTTTGTG GGATTTTCAA TTATTAATCT  
ATAATTTATT TTTACCAAAA TTTACAATCC CAAAAACAT CCTAAAAGTT AATAATTAGA

5650

CTATAATTCG ATGAACTAAG TAAAAAGCA TCAAACTTTC TTGGCAGAA CACATTTTTC  
GATATTAAGC TACTTGATTG ATTTTTTTCGT AGTTTGAAAG AACCGTCTTA GTGTAAAAAG

5700

5750

TCTAACTAA ATATGGACTG AAATGAAAA ATTAAACCAC TAGCTAGAAT AAAGTGTTGG  
AGATTTGATT TATACCTGAC TTTAACTTTT TAATTGGTG ATCGATCTTA TTTACAACC

5800

TGAGAGTGGA ACTCTAATTT CTCTCCTTTA CTAATTATGT ATAAACACAA AAATGCACCA  
ACTCTCACCT TGAGATTAAA GAGAGGAAAT GATTAATACA TATTTGTGTT TTTACGTGGT

5850

AATTTTTAGG TTTGAAAATA TCTAAGCATG GATAGGGTAA TTAACATTTT TTCTTTCAAT  
TTAAAAATCC AAACTTTAT AGATTCGTAC CTATCCCATT AATTGTAAAA AAGAAAGTTA

5900

TTTGCAATAT TTGAATAAAT CCTATGAGGG TCTTTGGTAC ACAATAATTG GAGGGTATAT  
AAACGTTATA AACTTATTTA GGATACTCCC AGAAACCATG TGTTATTAAC CTCCCATATA

5950

AGTTGAGTCT GAGAGTATAT TAGAAAAGAG ATATTTCAAG TAATGAAGCT GACATGTTTA  
TCAACTCAGA CTCTCATATA ATCTTCTCT TATAAAGTTC ATTACTTCGA CTGTACAAAT

6000

6050

TATGTACTTT GAGAGAAGTG TTGTGAGATT TGTACAAATG TATATGTACA CTTTAAAAAG  
ATACATGAAA CTCTCTTCAC AACACTCTAA ACATGTTTAC ATATACATGT GAAATTTTTC

6100

CAATATAAGA TAGATAAAAA AAATATAAAG AAAAAAGAA AGAAAGAAAG AAAGAAAGAG  
GTTATATTCT ATCTATTTT TTTATATTT TTTTTTCTT TCTTCTTTC TTTCTTTCTC

6150

AGAGGCTCAT ATATATATAG AATTGCTTGC AAGGAAAGAG AGAGAGAGAG ATTGAGATAT  
TCTCCGAGTA TATATATATC TTAACGAACG TTCCTTTCTC TCTCTCTCTC TAACCTCTATA

6200

CTTTTGGGAG AGGAGAAAAG AAAAGAAAAT GGGAGAGGG AGAGTAGAAT TGAAGAGGAT  
GAAAACCTC TCCTCTTCT TTTCTTTTA CCCTTCTCC TCTCATCTTA ACTTCTCCTA

6250

AGAGAACAAG ATCAATAGGC AAGTGACGTT TGCAAAGAGA AGGAATGGTC TTTTGAAGAA  
TCTCTTGTT TAGTTATCCG TTCACTGCAA ACGTTTCTCT TCCTTACCAG AAAACTTCTT

6300

6350

AGCATACGAG CTTTCAGTTC TATGTGATGC AGAAGTTGCT CTCATCATCT TCTCAAATAG  
TCGTATGCTC GAAAGTCAAG ATACACTACG TCTTCAACGA GAGTAGTAGA AGAGTTTATC

6400

AGGAAAGCTG TACGAGTTTT GCAGTAGTTC GAGGTATATA TCTACTTTTG TATATATATT  
TCCTTTCGAC ATGCTCAAAA CGTCATCAAG CTCCATATAT AGATGAAAAC ATATATATAA

6450

ACTTATAACA TAAACATTTT ATATACATAT TAAGTAACAC AAAAAATGCT TGTATGTATG  
TGAATATTGT ATTTGTAAAA TATATGTATA ATTCATTGTG TTTTACAGA ACATACATAC

Fig. 3g

09869582-022002

6500  
GGTCTCTCTG TGATGTGTTG TTGTGTCGTA CGTACGTGTT CTATCATATC CTTTAAAG  
CCAGAGAGAC ACTACACAAC AACACAGCAT GCATGCACAA GATAGTATAG GAAAATTTTC

6550 6600  
AAGCAAAGAG GAAAAAAAT TTGGGATACC CCAAATCTGT ATCATTTTAT AACAAAGTTTG  
TTCGTTTCTC CTTTTTTTA AACCCATGAG GGTTTAGACA TAGTAAATA TTGTTCAAC

6650  
CTTTTTGAT GTTCTTTGT GTTCTCTTT GATTTCATT TTTGTTTTG ATTTTTTTC  
GAAAAACTA CAAGAAAACA CAAAGAGAAA CTAAAGGTAA AAACAAAAC TAAAAAAG

6700  
TATTTCTCTT TACATCTATC AAAGTTTTT TTCTTATATT TTATTGCTTA TTTGTTTGT  
ATAAGAGAA ATGTAGATAG TTTCAAAAA AAGAATATAA AATAACGAAT AAACAAACAG

6750  
TACTTAATTC ACATTATCTG AGAGAAGAAC AATCTATCTG ATATGAAATT AGGGTTAATT  
ATGAATTAAG TGTAATAGAC TCTCTTCTTG TTAGATAGAC TATACTTTAA TCCCAATTAA

6800  
TCTCTTGTGA GTACTCTTTA ATTCACATAA GCTTAAAGTT TCCACCTTTT GATTCTGGGG  
AGAGAACACT CATGAGAAAT TAAGTGATTT CGAATTTCAA AGGTGGAAA CTAAAGACCC

6850 6900  
GTCGTCCAAT TCGATCAAAT CACTCAATTT TGTGTCAGA TTGATATAAG TTCATAGGGG  
CAGCAGGTTA AGCTAGTTTA GTGAGTTAAA ACAACAGTCT AACTATATTC AAGTATCCCC

6950  
GATATTGTTT CCACGACAAT CCATTTTAGT AACCCTAGG GGTTCCTCAAT TTTGGGTTTT  
CTATAACAAA GGTGCTGTTA GGTAAATCA TTGGGAATCC CCAAAGGTTA AAACCCAAAA

7000  
GAATTGACGC TAATGTCAA TTCATCTAAA GTCCGTGGA TATGTATACT TGGGGATGGG  
CTTAAGTCCG ATTACAGTTT AAGTAGATT CAGGCAACCT ATACATATGA ACCCTACCC

7050  
ATTCATCCTT TTTCTGGGT TCTTTAGATC TTCTCTTAA AACTAACAG ATTTTGTGT  
TAAGTAGGAA AAAAGACCCA AGAAATCTAG AAGAGAATTT TCTGATTGTC TAAAAACA

7100  
AAACCCTAGG AAACAGTTAA AAATCCCAT TTTAAAAACA TGTTTTGAAC TTGATGAGTA  
TTTGGGATCC TTTGTCAATT TTTAGGGTAA AAATTTTGT ACAAACCTTG AACTACTCAT

7150 7200  
AGATTAAATGG AAGAAATGAT GTTTTTGTGT GGTGTGAAGC ATGCTTCGGA CACTGGAGAG  
TCTAATTACC TTCTTTACTA CAAAAACACA CCACACTTCG TACGAAGCCT GTGACCTCTC

7250  
GTACCAAAG TGTAATATG GAGCACCAGA ACCCAATGTG CCTTCAAGAG AGGCCTTAGC  
CATGGTTTTT ACATTGATAC CTCGTGGTCT TGGGTACAC GGAAGTTCTC TCCGAATCG

7300  
AGTTGTACCC AATTCTCTTC TCTTTCTTCT AATTACCTTA ATTAATTACT CTCAATTTTT  
TCAACATGGG TTAAGAGAAG AGAAAGAAGA TTAATGGAAT TAATTAATGA GAGTTAAAA

7350  
ACTTTGATTT TTAGAGTCAA ATGATTAATG TTATAATTTG TCATATACTT CAGGAACCTA  
TGAAACTAAA AATCTCAGTT TACTAATTAC AATATTAAAC AGTATATGAA GTCCTTGAAT

7400  
GTAGCCAGCA GGAGTATCTC AAGCTTAAGG AGCGTTATGA CGCCTTACAG AGAACCCAAA

Fig. 3h

09869582-0022002

CATCGGTCGT CCTCATAGAG TTCGAATTCC TCGCAATACT GCGGAATGTC TCTTGGGTTT  
 7450 7500  
 GGTAAACTAA TTAGCTTCTT CAGCTACCTT CAGAGAGTGT TTGTTTTTTT AGTAGATTTT  
 CCATTTGATT AATCGAAGAA GTCGATGGAA GTCTCTCACA AACAAAAAAA TCATCTAAAA  
 7550  
 TTTGATGGTT TTGATGTTGA AATAGGAATC TGTTGGGAGA AGATCTTGGA CCTCTAAGTA  
 AAACCTACCA AACTACAACCT TTATCCTTAG ACAACCCTCT TCTAGAACCT GGAGATTCAT  
 7600  
 CAAAGGAGCT TGAGTCACTT GAGAGACAGC TTGATTCTTC CTTGAAGCAG ATCAGAGCTC  
 GTTTCCTCGA ACTCAGTGAA CTCTCTGTCT AACTAAGAAG GAACTTCGTC TAGTCTCGAG  
 7650  
 TCAGGGTACT ACTTTGTTCA TCAATATCTT TATACACTGA TCTATTTCCA TAGTAAGATT  
 AGTCCCATGA TGAACAAGT AGTTATAGAA ATATGTGACT AGATAAAGGT ATCATTCTAA  
 7700  
 AAATTTGGTG TTTAATTCTG CAGACACAGT TTATGCTTGA CCAGCTCAAC GATCTTCAGA  
 TTTAAACCAC AAATTAAGAC GTCTGTGTCA AATACGAACCT GGTCGAGTTG CTAGAAGTCT  
 7750 7800  
 GTAAGGTAAA TAAAGAAACA CTCATTCTCC TCTCTAAATT CCTCATCTAA AAGTAATGTA  
 CATTCCATTT ATTCTTTTGT GAGTAAGAGG AGAGATTTAA GGAGTAGATT TTCATTACAT  
 7850  
 ACCAAGAAAA CACAAATATT TGGAGCAGGA ACGCATGCTG ACTGAGACAA ATAAAACTCT  
 TGGTTCCTTT GTGTTTATAA ACCTCGTCTT TCGTACGAC TGACTCTGTT TATTTTGAGA  
 7900  
 AAGACTAAGG GTAATTAATA TACATTCTCA TATCACCAAA TTAATGCATC ACTAAATTTG  
 TTCTGATTCC CATTAATTAT ATGTAAGAGT ATAGTGGTTT AATTACGTAG TGATTTAAAC  
 7950  
 GTTATAATGT GTGTGTGTAT ATACATATGT GACAGTTAGC TGATGGGTAT CAGATGCCAC  
 CAATATTACA CACACACATA TATGTATACA CTGTCAATCG ACTACCCATA GTCTACGGTG  
 8000  
 TCCAGCTGAA CCCTAACCAA GAAGAGGTTG ATCACTACGG TCGTCATCAT CATCAACAAC  
 AGGTCGACTT GGGATTGGTT CTCTCTCAAC TAGTGATGCC AGCAGTAGTA GTAGTTGTTG  
 8050 8100  
 AACAACTCTC CCAAGCTTTC TTCCAGCCTT TGAATGTGA ACCCATCTT CAGATCGGGT  
 TTGTTGTGAG GGTTGAAAG AAGGTCGGAA ACCTTACACT TGGGTAAGAA GTCTAGCCCA  
 8150  
 AACTTTAGAC TAGTATAACC AATTTGATTT GAGTTCTATT ATAAGCTTTT CTTAAGAAAG  
 TTGAAATCTG ATCATATTGG TTAAACTAAA CTCAAGATAA TATTCGAAAA GAATTCCTTC  
 8200  
 TATCTCAAAC TACTAAATTT TATGGAGCAG GTATCAGGGG CAACAAGATG GAATGGGAGC  
 ATAGAGTTTG ATGATTTAA ATACCTCGTC CATAGTCCCC GTTGTCTAC CTTACCTCTG  
 8250  
 AGGACCAAGT GTGAATAATT ACATGTTGGG TTGGTTACCT TATGACACCA ACTCTATTTG  
 TCCTGGTTCA CACTTATTAA TGTACAACCC AACCAATGGA ATACTGTGGT TGAGATAAAC  
 8300  
 AATCTTTCTC ACTTAATCAA TCCCTCTCTT TTTTTTTTGA CATTTTAAAG ATGATGTTTC  
 TTAGAAAGAG TGAATTAGTT AGGGAGAGAA AAAAAAACT GTAAAAATTC TACTACAAAG

Fig. 3i



8350  
TATTTTATTA CCTCTCTCAT GTTTTCTGTC TTGTGTGCAT GTGTGTGTGT AATGTTTATG  
ATAAAATAAT GGAGAGAGTA CAAAAGACAG AACACACGTA CACACACACA TTACAAATAC

8400  
CCCTTCTATT ATTCAATAAT TTTTTCGACA ATTTTGCTTC CTATTTTATC CCATTACTCC  
GGGAAGATAA TAAGTTATTA AAAAAGCTGT TAAACGAAG GATAAAAATG GGTAAATGAGG

8450  
TAAACTTCCT GATCCAGTTT CTTTAAAT AACTCCCATT TTATGCATGT TATCTAACCA  
ATTTGAAGGA CTAGGTCAA GAAATTTTA TTGAGGGTAA AATACGTACA ATAGATTGGT

8500  
ATTCTCTTAA CTATGATTTA TGGTACGATA TAACTCACAG TCTCACACTA TCTATTTGGT  
TAAGAGAATT GATACTAAAT ACCATGCTAT ATTGAGTGTC AGAGTGTGAT AGATAAACCA

8550  
TTTTTTTTGT TTGAGTCTTG AGAAGGGACC GCTTGTTTAT CTCTCTTGTT AAAGAGCAAC  
CAAAAAACA AACTCAGAAC TCTTCCCTGG CGAACAAATA GAGAGAACAA TTTCTCGTTG

8600  
TCACTGGCCA CTGCTTATGT ATCTGTAGGC CCCACCTATA TCATTTTGGC TATATCTATA  
AGTGACCGGT GACGAATACA TAGACATCCG GGGTGGATAT AGTAAAACCG ATATAGATAT

8650  
CTTTTGTAGA GGGAGTATTA CTATAGAGAA GAAGATAAAT TTGGTTCTAA TATATCTTGC  
GAAAACATCT CCCTCATAAT GATATCTCTT CTTCTATTTA AACCAAGATT ATATAGAACC

8700  
AGGTAGTTGA TATTCTCAAT TATCATGAAG ATTTGATAGA CAAGTTTATC AGATACCTTA  
TCCATCAACT ATAAGAGTTA ATAGTACTTC TAACTATCT GTTCAAATAG TCTATGGAAT

8750  
AACATAGGTT TAAGATCTCA ATTGAAATGT GAATTCACCC GACGATTAGA GTTACGATCT  
TTGTATCCAA ATTCTAGAGT TAACTTTACA CTTAAGTGGG CTGCTAATCT CAATGCTAGA

8800  
AAGGAAGCGT TTCTTGAATT TTGAGTTTGT TTGATCAAGA GTAGAATGCT TTTCTATTAC  
TTCCTTCGCA AAGAAGTTAA AACTCAAACA AACTAGTTCT CATCTTACGA AAAGATAATG

8850  
TAAGGTTGTT AATGCTTATA TTCCATGACC AAGGCCAAGA GAACAAACAA AAACATGGTG  
ATTCCAACAA TTACGAATAT AAGGTACTGG TTCCGGTTCT CTGTTTGTGTT TTTGTACCAC

8900  
CCTCTTGATG TATAGTAATG GCTCTTAATG GTCATATACA GAGAAAAAAA GATTAATGTC  
GGAGAACTAC ATATCATTAC CGAGAATTAC CAGTATATGT CTCTTTTTTT CTAATTACAG

9000  
GTTGCACAAG CTTGAAGTTA CTTACTCCTC GTCTTCCTCA TTAGTGTCTT CGTCTTCCTC  
CAACGTGTTT GAACTTCAAT GAATGAGGAG CAGAAGGAGT AATCACAGAA GCAGAAGGAG

9050  
ATCCTCATCG CTCCCAATAT AGGGCTTCAT CTAATTGAAA ACCAAATGCT CATGCAGTGG  
TAGGAGTAGC GAGGGTTATA TCCCGAAGTA GATGAACCTT TGGTTTACGA GTACGTCACC

9100  
AAAAAGATAA CAGAGGTTCA AATTAAGGCA AACAAACTA CAAGTGAGAA AGGGAAACTA  
TTTTTCTATT GTCTCCAAGT TTAATTCCGT TTGTTTGTGAT GTTCACTCTT TCCCTTTGAT

9150  
CAAGTGGTAA GATGTAATGT TTTGACTCAA AACCAGATCA GACAATGAAA AAAAGTATTG

9200  
CAAGTGGTAA GATGTAATGT TTTGACTCAA AACCAGATCA GACAATGAAA AAAAGTATTG

9250  
CAAGTGGTAA GATGTAATGT TTTGACTCAA AACCAGATCA GACAATGAAA AAAAGTATTG

9300  
CAAGTGGTAA GATGTAATGT TTTGACTCAA AACCAGATCA GACAATGAAA AAAAGTATTG

Fig. 3j

GTTCACCATT CTACATTACA AACTGAGTT TTGGTCTAGT CTGTTACTTT TTTTCATAAC  
 ATACAAAAAG TCCATCCGGA AGCATAATTA CCGCTTGCGAG GATGTCATCA GAGATGTCTG  
 TATGTTTTTC AGGTAGGCCT TCGTATTAAT GGCGAACGTC CTACAGTAGT CTCTACAGAC  
 9350  
 TAGTTCGGCC AATGGCATAG ATGGTGAGCG GACCAGAGTA GCGTAAATCC TCTAAATACT  
 AATCAGCCGG TTACCGTATC TACCACTCGC CTGGTCTCAT CGCATTTAGG AGATTTATGA  
 9400  
 GTCTAAAAGC CGGACCGACC CGACAAGGAT CACAGTCAAG GGAATAGGA CACCTATTGA  
 CAGATTTTCG GCCTGGCTGG GCTGTTCTTA GTGTCAGTTC CCCTTATCCT GTGGATAACT  
 9450  
 TATCCCAAAA GACTGTTGTT ACAGCCACAT CATCCTTGTC CAACTGGGTA GCCCAAAGGG  
 ATAGGGTTTT CTGACAACAA TGTGCGTGTA GTAGGAACAG GTTGACCCAT CGGGTTTCCC  
 9500  
 AACTAGTTG TGSTAAGAGC TTGTTTGACT CAAAAAATGG CTAAGTAGGA TGATGCTGAA  
 TTTGATCAAC ACCATTCTCG AACAAACTGA GTTTTTTACC GATTGATCCT ACTACGACTT  
 9550  
 TTACCATCTG TTCATGTTTT TGAAGTAGA GATGGGTAGT GAAATTTTCA AAGCCTTTGC  
 AATGGTAGAC AAGTACAAAA ACTGATCTCT CTACCCATCA CTTTAAAGT TTCGGAAACG  
 9600  
 AAAACGCCTG TGGGACCTGT TTCAGAAAAA GACTTAAAAG ACTTGAGACT CAAGGAAAAT  
 TTTTGCGGAC ACCCTGGACA AAGTCTTTTT CTGAATTTTC TGAAGTCTGA GTTCCTTTTA  
 9650  
 AATATCCATT ATATAAAGAT GACAACAAAT ATTAACGGAA GTAGGAGTGA TTGAGAACGA  
 TTATAGGTAA TATATTTCTA CTGTTGTTTA TAATTGCCTT CATCCTCACT AACTCTTGCT  
 9700  
 TTCTAGTAGA AGAGACGGCT CGCAGGACGT CGTTTATAAT AGGCCAATGG CAGAGATAGT  
 AAGATCATCT TCTCTGCCGA GCGTCTGCA GCAAATATTA TCCGGTTACC GTCTCTATCA  
 9750  
 GAGAGGACCG GAGTAGCCTA AATTCTTTAA ATGTCGTTTG ATACACGGAC CAACTAGACG  
 CTCTCCTGGC CTCATCGGAT TTAAGAAATT TACAGCAAAC TATGTGCCTG GTTGATCTGC  
 9800  
 AGCATCATAC TCAGAGGGAA CCGGACACGT CTTGATATCC CAGAAGACCG ATGTTACGGC  
 TCGTAGTATG AGTCTCCCTT GGCCTGTGCA GAACTATAGG GTCTTCTGGC TACAATGCCG  
 9850  
 CTTAGCTTGC TGCCGCGTTG CCTTCATCAT CATCTTCTCC TTTTAATCTA TAACGGAAAT  
 GAATCGAACG ACGGCGCAAC GGAAGTAGTA GTAGAAGAGG AAAATTAGAT ATTGCCTTTA  
 9900  
 CAAACATCAG ATAAAGCATT CGAAAAGATA GATTGACACA GGTAAATCA TCCACTTCAG  
 GTTTGTAGTC TATTTCTGTA GCTTTTCTAT CTAAGTGTGT CCAATTTAGT AGGTGAAGTC  
 9950  
 AGAAAAAGAG AGGGACATGG CCGTAAACAA TGAGATAAGG ATCGGCCTAA TGTTTATAAT  
 TCTTTTCTC TCCCTGTACC GGCATTTGTT ACTCTATTCC TAGCCGATT ACAAATATTA  
 10000  
 GGGCTTGCGT TTAATGGGCC TACAGTTTCT TGAATCAGCC TTATGCATGA GTCCTAGTAT  
 CCCGAACGCA AATTACCCGG ATGTCAAAGA ACTTAGTCGG AATACGTACT CAGGATCATA  
 10050  
 10100  
 10150  
 10200

Fig. 3k

10250  
TTTATCAACT TTTTTTTTTC ATCTTTCTTT AGTTACAATA GATTTAAAGT GTTTTTTGT  
AAATAGTTGA AAAAAAAG TAGAAAGAA TCAATGTTAT CTAAATTTCA CAAAAACAA

10300  
AATGCCATTG CAAAATTTGG TAACTGTTTA TAACATTGTT CCTCACTTCA AAATTTAAAG  
TTACGGTAAC GTTTTAAACC ATTGACAAAT ATTGTACAA GGAGTGAAGT TTTAAATTTG

10350  
CACCATTAAT AAAAGCTATA CATATAATTA TAACTTGGGT TTTGTGCAA AAAACAAAC  
GTGGTAATTA TTTTCGATAT GTATATTAAT ATTGAACCCA AAACACGTTT TTTTGTGTTG

10400  
AAATTAACCT TTCATTTTAA ATAAATGCAA TTCAATACCG CAATATCAAA AGTAACCCGT  
TTAATTGGA AAGTAAATTT TATTACGTT AAGTTATGGC GTTATAGTTT TCATTGGGCA

10450  
ATAACCTTTA TTCGTGTATA GATTTTAGAA ACAGTATAAG TCAAATTATC AAAACTATGT  
TATTGGAAAT AAGCACATAT CTAAAATCTT TGTCAATTC AGTTTAATAG TTTTGATACA

10500  
TGTTTTAAGC ATTTTAAAA TAAGAATAAT AATAATGTTG AAGGGTGGAT TTGAACCCAT  
ACAAAATTCG TAAATTTTTT ATTCTTATTA TTATTACAAC TTCCACCTA AACTTGGGTA

10550  
GAACTATAGA ACAAACCAAA GCATGCATAA CCACATGCGC CGAACAAACC AAAAATCAT  
CTTGATATCT TGTTTGGTTT CGTACGTATT GGTGTACGCG GCTTGTGTTG TTTTGTGAGTA

10600  
GGCTTTGTTA AACATATAAA AATATTCGAA TAAAAAATGT GGGGAACCTG TTACCAGTTT  
CCGAAACAAT TTGTATATTT TTATAAGCTT ATTTTTTACA CCCCTTGAAC AATGGTCAAA

10650  
TGGTTCTTTT TGGAGCCATT TTTTCAACA CAGATATTGT TAAGGAGTTT CAGGTAAAC  
ACCAAGAAAA ACCTCGGTAA AAAAAGTTGT GTCTATAACA ATTCTCAAA GTCCATTTTG

10700  
TGTATATTAT GCAGGGAACC ACAGTAGGCT ATAATGAAAG TCACACTGTG AAGTTAGCAG  
ACATATAATA CGTCCCTTGG TGTACATCCG TATTACTTTC AGTGTGACAC TTCAATCGTC

10750  
ACAAGTTTTT ACTTAAAGAT GTGAGTTGTG ATCTTTTGA TGTAAGTCTT GATGTATATG  
TGTTCAAAAA TGAATTTCTA CACTCAACAC TAGAAAACT ACATTCAGAA CTACATATAC

10800  
TTGACAAAT ATATAAGTTT GTATTGCATA TTCTATGACT TACGAAGTTT CTATGCAAGA  
AACTGTTTAA TATATTCAAA CATAACGTAT AAGATACTGA ATGCTTCAAA GATACGTTCT

10850  
AAAGCCGGGA GAAATTTTCC GTCAAGTAAC TAAGAGATCG TAATCTTGT CTGAAGAACA  
TTTCGGCCCT CTTTAAAGG CAGTTCATTG ATTCTCTAGC ATTAAGAACA GACTTCTTGT

10900  
ACCTTTTTTT ATTATTTGAG TTTAGGTTGC CAACAGTGAA CAAAGGGACG AGATACCATA  
TGGGAAAAAA TAATAAATC AAATCCAACG GTTGTCACTT GTTCCCTGCT TCTATGGTAT

10950  
TGACAAATAT CCTCTAACGC CATTTCAACA GTTAATCAAC AGTGTCCGCT ATATGCATGT  
ACTGTTTATA GGAGATTGCG GTAAAGTTGT CAATTAGTTG TCACAGCCGA TATACGTACA

11000  
GCTAACAAATG CACAAGAACA TTGTCACCAT CCCGTGAATA TGAATATTAA TGATTATGAA

11050  
GCTAACAAATG CACAAGAACA TTGTCACCAT CCCGTGAATA TGAATATTAA TGATTATGAA

11100  
GCTAACAAATG CACAAGAACA TTGTCACCAT CCCGTGAATA TGAATATTAA TGATTATGAA

11150  
GCTAACAAATG CACAAGAACA TTGTCACCAT CCCGTGAATA TGAATATTAA TGATTATGAA

Fig. 3I

CGATTGTTAC GTGTTCTTGT AACAGTGGTA GGGCACTTAT ACTTATAATT ACTAATACTT

11200  
CGAGTTTGTA GAGTTCCAAG AGGAAGGTAC TACCTTCTCA TACTCATTGA TCATATATTT  
GCTCAAACAT CTCAAGGTTT TCCTTCCATG ATGGAAGAGT ATGAGTAACT AGTATATAAA

11250  
TGTTTCTTGT TTGTTTTAGT AACTAGGGTT ATTCCGATTG TTTTTCAAAA TAATAGTAAT  
ACAAAGAACA AACAAAATCA TTGATCCCAA TAAGCCTAAC AAAAAGTTTT ATTATCATT

11300  
ATGTCAACTA TATTTATAAA AAAAAAACT AAATAACTTT TGTACAATTG ATCATTTTTT  
TACAGTTGAT ATAAATATTT TTTTTTTTGA TTTATTGAAA ACATGTTAAC TAGTAAAAAA

11350 11400  
AAATATATCA TAAAGATTCA TCAATATATG AACATATATT TTTAACAAAT AACTAATTG  
TTTATATAGT ATTTCTAAGT AGTTATATAC TTGTATATAA AAATTGTTAA TGTGATTAA

11450  
GCTATATAGT GTATAGTTCC TTTTGTGGAG AGGTTTAACT TCAGTTCAGA GATTATTGTA  
CGATATATCA CATATCAAGG AAAACACCTC TCCAAATTCA AGTCAAGTCT CTAATAACAT

11500  
CTTGGTAAAA TATTTGTCCT TGTTAATTAG TTCATCTTCT AGAATACAGA TTTGGGCCAT  
GAACCATTTT ATAAACAGGA ACAATTAATC AAGTAGAAGA TCTTATGTCT AAACCCGGTA

11550  
GTAGTTTCCC AGAAAACACC GGAAAAAAA TTCACACTTC ACACCAGAAA CAATAAACGA  
CATCAAAGGG TCTTTTGTGG CCTTTTTTTT AAGTGTGAAG TGTGGTCTTT GTTATTTGCT

11600  
GGAACAGAGC CCAAACATCAT CCCTATAATT GGGCCCCAAA AAAGCAGAGC AAACCAAACC  
CCTGTCTCTG GGTGTGAGTA GGGATATTAA CCCGGGTTTT TTTCGTCTCG TTTGGTTTGG

11650 11700  
AAAATCAAGT AAATCCATTT ACAAATATGC TTTATAATTA TTATTTTCT CAACCACAA  
TTTTAGTTCA TTTAGGTAAA TGTTTATACG AAATATTAAT AATAAAAAGA GTTGGTGT

11750  
TATGCTTTAT AATTTATGTA AATGTTATAT GAATTATTTA CGATTATTT TAATTACTTT  
ATACGAAATA TTAAATACAT TTACAATATA CTTAATAAAT GCTAAATAAA ATTAATGAA

11800  
ATCTTGAAT TATCTTACGA AGTTAATGAA AATATTTTAA ATATCTAATT TATATATGTC  
TAGAACCTTA ATAGAATGCT TCAATTACTT TTATAAAATT TATAGATTAA ATATATACAG

11850  
TGGACTAAAA TAAATAGAAA TATCTGTATT CCAATCATCA CAAAAAAA ATTCTCATCA  
ACCTGATTTT ATTTATCTTT ATAGACATAA GGTTAGTAGT GTTTTTTTTT TAAGAGTAGT

11900  
TCTTTGATAT ATAGAAAGTT TTTAAATTT CAGTTTCACA GATTTTACCA ATTATAGTTT  
AGAACTATA TATCTTTTCAA AAATTTTAAA GTCAAAGTGT CTAAATGGT TAATATCAAA

11950 12000  
TATAAGCTTA TGCTAATTAT GTGATCAATG CAAACAAAAG TTGACAATAA TAAATGAAG  
ATATTGGAAT ACGATTAATA CACTAGTTAC GTTTGTTTTC AACTGTTATT ATTTTACTTC

12050  
TCAAATATGA TAGATTCCTA CTATAAATAT AGACTCGTGA ATAATACTCG AATCAGTCTC  
AGTTTATACT ATCTAAGGAT GATATTTATA TCTGAGCACT TATTATGAGC TTAGTCAGAG

Fig. 3m

12100  
TGAGGTTTTG CTGGAAGA AAAACCGAAG AGCTCAAAAC AGAGTGCCTT TGTTCCTGGG  
ACTCCAAAAC GACCTTTTCT TTTTGGCTTC TCGAGTTTTG TCTCAGCAA ACAAGACCC

12150  
AATCTTCAAG CCTCTCACTT GCGAAGACGA AGCTTACTCG TAAGGTGATT ATCTTCTTCT  
TTAGAAGTTC GGAGAGTGAA CGCTTCTGCT TCGAATGAGC ATTCCACTAA TAGAAGAAGA

12200  
TCTTCTTCTT TTCAATTCTT TTTTCGTTCA TCTGAAATGT GAAATCATGT GACGTGACGA  
AGAAGAAGAA AAGTTAAGGA AAAAGCAAGT AGACTTTACA CTTTAGTACA CTGCACTGCT

12250 12300  
TTAGGTAAAC GATCGAATTT CTTAATTTCT TATATGATTA TCTTCTAGTT TCTTGATCAG  
AATCCAATTG CTAGCTTAAA GAATTAAAGC ATATACTAAT AGAAGATCAA AGAACTAGTC

12350  
CACATCTTGT TGTTCCTTTT CAATCGAGAC TGATTCTAGA TGTTCCTAAG GATCTTGTTC  
GTGTAGAACA ACAAAGAAA GTTAGCTCTG ACTAAGATCT ACAAGAATTC CTAGAACAAG

12400  
GATGAACTTT GCATGAATCA TCCATATCGA CGAACTGGTC TGATCTTCTT GTTGTATATGG  
CTACTTGAAA CGTACTTAGT AGGTATAGCT GCTTGACCAG ACTAGAAGAA CAACAATACC

12450  
ATTAAGTTTC TTGAGATACA AGAAAGGCTT CAATGATCAA TCTGATCTGT TTTGATGAAC  
TAATTCAAAG AACTCTATGT TCTTCCGAA GTTACTAGTT AGACTAGACA AACTACTTG

12500  
ACAAATCTTT ATCTTTGAAC CATGGATAAG GTCAATTTCA CACCATGGCT GGAGGAAGTT  
TGTTTAGAAA TAGAACTTG GTACCTATTG CAGTTAAAGT GTGGTACCGA CCTCCTTCAA

12550 12600  
TATCACCGGC GTCATCTTTG GAAGATGTAA AGGCATACGT CAATGCTGTG GAGGTGCGAT  
ATAGTGGCCG CAGTAGAAAC CTTCTACATT TCCGTATGCA GTTACGACAC CTCCAGCGTA

12650  
TGCAGGAAAT GGAACCTGCA AGATTGGAA TGTGTGTAAG ACTCTTTCGT GGTTTACAG  
ACGTCCTTTA CCTTGGACGT TCTAAACCTT ACAAACATTC TGAGAAAGCA CCAAATGTG

12700  
CTCCTAGGTG TGTGTGTTT GCTCTTAAAC AGTCTAAAGA ACAATGACAC ATGTGAGAAT  
GAGGATCCAC ACAAACCAA CGAGAATTG TCAGATTCT TGTACTGTG TAACTCTTA

12750  
TGATTCTGAT GTTATTTTTC TCTTGTAGG ATCGGTATGC CTACTTTCAG TGCACGCATG  
ACTAAGACTA CAATAAAAAG AGAAACATCC TAGCCATACG GATGAAAGTC ACGTGCCTAC

12800  
CAGGACCTCT TGAAAGATCA CCCGAGTCTG TGTCTTGGTT TAAATGTCTT ACTTCCACCT  
GTCCTGGAGA ACTTTCTAGT GGGCTCAGAC ACAGAACCAA ATTTACAGAA TGAAGGTGGA

12850 12900  
GAGTATCAGT TAACCATACC TCCCAGGCT AGCGAAGAGT TTCATAAGGT GGTGGAAGA  
CTCATAGTCA ATTGGTATGG AGGCTCCGA TCGCTTCTCA AAGTATCCA CCAACCTTCT

12950  
AGCGTACCAG TACCACCAA GGTGGTTGGA AGAAGTCTAC CACGTCCGGA GCCTACCATA  
TCGCATGGTC ATGGTGGTTT CCACCAACCT TCTTCAGATG GTGCAGGCCT CGGATGGTAT

13000  
GATGATGCGA CTTCATACCT TATTGCTGTG AAGGAAGCCT TTCATGATGA ACCTGCAAAA

Fig. 3n

CTACTACGCT GAAGTATGGA ATAACGACAC TTCCTTCGGA AAGTACTACT TGGACGTTTT

13050  
TATGGGGAAA TGCTTAAGCT CTTGAAAGAT TTAAAGCTC GCAGGTATGT ATTAGTTCTT  
ATACCCCTTT ACGAATTCGA GAACCTTCTA AAATTTTCGAG CGTCCATACA TAATCAAGAA

13100  
TTCTCCATGT TATGTTTGAT TTTTTCAGTC TACAGAACAA ACACATTATG TGAATTGATT  
AAGAGGTACA ATACAAACTA AAAAAGTCAG ATGTCTTGTT TGTGTAATAC ACTTAACATA

13150 13200  
CTGATGTTAC TAAGTCTCTT TGTAGAGTCG ATGCCGCTTG TGTCATTGCT AGGGTGGAGG  
GACTACAATG ATTCAGAGAA ACATCTCAGC TACGGCGAAC ACAGTAACGA TCCCACCTCC

13250  
AACTCATGAA AGATCACTTG AATCTGCTTT TTGGTTTCTG TGTCTTCCTT TCAGCTACAA  
TTGAGTACTT TCTAGTGAAC TTAGACGAAA AACCAAAGAC ACAGAAGGAA AGTCGATGTT

13300  
CGAGTTTTAC CACGAAGCTT AAGGTATAGA GTGCTTATAG TTACCATTG ATGTTTCCTA  
GCTCAAAATG GTGCTTCGAA TTCCATATCT CACGAATATC AATGGTAAAC TACAAAGGAT

13350  
TATGTTAACT TGTGGTTTAA GTACAAAAT TGTCCATGTG CAGGCAAGGT TTCAGGGCGA  
ATACAATTGA ACACCAAATT CATTGTTTTA ACAGGTACAC GTCCGTTCCA AAGTCCCGCT

13400  
TGGTAGTCAA GTAGTTGACT CAGTTCTTCA GATAATGAGA ATGTACGGTG AGGGAAACAA  
ACCATCAGTT CATCAACTGA GTCAAGAAGT CTATTACTCT TACATGCCAC TCCCTTTGTT

13450 13500  
GTCCAAACAT GATGCGTATC AGGAGGTAGG CTTCTTGCTA GGATACTTTG TGTGTGTGT  
CAGGTTTGTG CTACGCATAG TCCTCCATCC GAAGAACCAT CCTATGAAAC ACAACACACA

13550  
TGCACCTTCT TAGTTCTTTG GTTTGATTG CTTTGTATC TTTTGCAAGT CGTTGCACTT  
ACGTGAAAGA ATCAAGAAAC CAACTAAAC GAAACAATAG AAAACGTCCA GCAACGTGAA

13600  
GTTTCAGGTC ATGACGATTT AGTCATGGAG CTTTCACAAA TTTTGAAGTGA TCCACCTACT  
CAAGTCCCAG TACTGCTAAA TCAGTACCTC GAAAGTGTTT AAAACTGACT AGGTGGATGA

13650  
GGAGTCTAGA GATAGCCAGA TAGCTAAGGA GAGTACTGGA AGACTGTAAT ATACCATAAG  
CCTCAGATCT CTATCGGTCT ATCGATTCTT CTCATGACCT TCTGACATTA TATGGTATTC

13700  
AGACGAAAAA GAAAGTAGAG CTTCTCACGA AAAGAGAGTG TTTTGTAGTT TCTTTTGCAA  
TCTGCTTTTT CTTTCATCTC GAAGAGTGCT TTTCTCTCAC AAAAATCAAA AGAAAACGTT

13750 13800  
ACATTAGAGT TTTGTTTGAT TAACATGACA TTCAAAAATA TGCTATGCTT CTATGTTGAG  
TGTAATCTCA AAACAAACTA ATTGTACTGT AAGTTTTTAT ACGATACGAA GATACAACCT

13850  
GTGTACAATG AATTGGTGTA TAAGAGACTA AAAGAGAGTG TATAGTTTCT TTGTTGAGGT  
CACATGTTAC TTAACCACAT ATTCTCTGAT TTTCTCTCAC ATATCAAAGA AACAACTCCA

13900  
TTCTTTTATG TTGAGGTGTT CAATATGCTA TTTTCAGGGT AATCTTTTTA TAAGAAACTG  
AAGAAAATAC AACTCCACAA GTTATACGAT AAAAGTCCCA TTAGAAAAAT ATTCTTTGAC

Fig. 3o

09869582-023807

27/43

13950  
AGAAGGGAAA CACTCAAAAA ACAGAGTTCA ACGTAGAAAC AAAAAACAGAG AGGTGAACCTG  
TCTTCCCTTT GTGAGTTTTT TGTCTCAAGT TGCATCTTTG TTTTGTCTC TCCACTTGAG

14000  
ATGAAAGATC AATTTAACCT GCTTGTGATG ATTGGCTTAT CAAGAGAATT GAAGAGATTG  
TACTTTCTAG TTAAATTGGA CGAACACTAC TAACCGAATA GTTCTCTTAA CTCTCTCTAAG

14050 14100  
ACGATTACAC AAATTCAATT CTAAAGACA AGAGTAGACT GCTAATTCTT ATTAAGGCTG  
TGCTAATGTG TTAAAGTTAA GAATTTCTGT TCTCATCTGA CGATTAAGAA TAATTCCGAC

14150  
TTAATGCTTC TTGAGAGCAT TGACCTTTTC CCTGAGGTAA TAAAGCTTGG CTCTTCTTAC  
AATTACGAAG AACTCTCGTA ACTGGAAAAG GGACTCCATT ATTTCAAGAC GAGAAGAATG

14200  
TTTCTTCTTG TCCACCACCT TAATCACCTT CAGGTTTGGG GAATACCTGT CACCAAAACA  
AAAGAAGAAC AGGTGGTGGA ATTAGTGGGA GTCCAAACCC CTTATGGACA GTGGTTTTGT

14250  
CCTCCACTTA CATCAGTATT TTCCATGACC AAGGCAAACA AAGAGAACAT ACAAACATG  
GGAGGTGAAT GTAGTCATAA AAGGTACTGG TTCCGTTTGT TTCTCTTGTA TGTTTTGTAC

14300  
GTGGCTCTTG ATTATAATAA TGGCTCTTAA TGGTCATATA CAAAAGTCTG AGAGAAAAG  
CACCGAGAAC TAATATTATT ACCGAGAATT ACCAGTATAT GTTTTCAGAC TCTCTTTTTT

14350 14400  
ATTAAAGTGG CTGCACAAGC TTGAAGCTTG AAGTTACTTA CAAGGGGAAC ATGGATTCTGA  
TAATTTTACC GACGTGTTCT AACTTCGAAC TTCAATGAAT GTTCCCTTG TACCTAAGCT

14450  
CGCCCACTCC AGCAACAAGC CTTCTAATTC TAAATGTTGA GTTGAGACCA GCATTACGCC  
GCGGGTGAGG TCGTTGTTCT GAAGATTAG ATTTACAAC CAACTCTGGT CGTAATGCGG

14500  
TTGCTATGAC GACGCCTTTT ACGATTGATA CACGCCTCTT GTTCTCAGGC ACTTCCTGTT  
AACGATACTG CTGCGGAAAA TGCTAACTAT GTGCGGAGAA CAAGAGTCCG TGAAGACAA

14550  
CAAAACAAGT AAATGAAAGG TTTCACCTAG AAGATGAAAG ATAGTTTGAT CTTACTCACC  
GTTTGTTCAT TTTACTTTCC AAAGTGAATC TTCTACTTTC TATCAAACTA GAATGAGTGG

14600  
CAAGAAAAAG AAATTACAAC CTAGGCCAAC AGTAGTTACC ACTTTTAGCT GCACAATGTA  
GTTCTTTTTT TTTAATGTTG GATCCGGTTG TCATCAATGG TGAAAATCGA CGTGTTACAT

14650 14700  
ACCAGGCTTT ATCTCTGGAA TCTCTCTAAG AGTTCTCACT TCCTCAACTG CTTCTTGTC  
TGCTCCGAAA TAGAGACCTT AGAGAGATT TCAAGAGTGA AGGAGTTGAC GAAGGAACAG

14750  
TACAATCTGC AGAGGATTGT GACATCGGTG CTTCTTGTC TACATGATAT ATCTAAATAC  
ATGTTAGACG TCTCCTAACA CTGTAGCCAC GAAGGAACAG ATGTACTATA TAGATTTATG

14800  
AAGTGTCAGG TTCGAGTTGT AGTACCTGCA TAATATGCTT AGCGGTTTTA TCAAGCCGCT  
TTCACAGTTC AAGCTCAACA TCATGGACGT ATTATACGAA TCGCCAAAAT AGTTGGGCGA

14850  
TAAACTTGAT TCTCTGAGGC ACAACACAAT CTGACTCAGG GGATCCTTGA ACAGAATCTC

Fig. 3p

09869582-022002

09/869582

WO 00/23578

PCT/US99/24407

28/43

ATTTGAACTA AGAGACTCCG TGTGTGTGTTA GACTGAGTCC CCTAGGAACT TGTCTTAGAG

14900

CAGTGGTGGG AAAACACCTC GACGAAAAGT TTTGTTTCTG CCAAAAAAAT ATTCCCAAGA  
GTCACCACCT TTTTGTGGAG CTGCTTTTCA AAACAAAGAC GGTTTTTTTA TAAGGGTTCT

Fig. 3q

09869582-022407



(2) CCCTCACACATTTCTTATCTTTTGCTCTCAATAGATTCCATTGATTCAAAAACAAAATTTTCATTAAAGATTTCACAACCTCCACACA 86  
 (4) -----GATTCA-CAAAAACCTTTTC-TTCAGATT-CACAATCTCATCAAA 42  
 (2) ---CTTCC-----AAACACAATTAAAGAGAGGAGGAAAAAGAAATCAATAACCTTATAAATAAAAAATCAGACAAAACAGA 154  
 (4) CCCTTCAAAAAGAGAAAAGATCTAAAGAATAAACAAAGAGCCCTAATATCAAAATCACAACCAAAAAAACCAAGAAAG-CTAATTAA 127  
 (2) AGTTTCTCTCTTCTTCTTCTTAAAGCTAGTACCTTTTCTCTTGAAA-TTAGGGTTAATTCTTTTTCAAAATACCATCAATTCT 238  
 (4) AGTTTCTCTCTAGCTATTCCTCTT---CTTTCTTCTCTTGAAAACCTAGGGTTTACTT----- 184  
 (2) CCAGACCATAAAAACTCAAAAAGATCAGATCTTTCTCTGAAAAAGAGATACCCCACTTATGTTTTTGTGTGTCTGTATATAG 321  
 (4) -----CACCAAAAGATAAGATCTTTTCCCCAGAAAAGCAATACCCAAGTCATGTTTCTGTGTGTCTGTATATAG 253  
 (2) ATAAA-CATTACATACCCATATTGTGTATAGACATAAAAAAGTGGAAATTAAGGTAACAAAAAGAA----- 386  
 (4) ATAAAACATTACATACCCCTAATAAGGTTACACAAATAGCTATAAAAGAGGGAAAAATAAGATAGGGATTTTGGGGGTAGGAAAG 338  
 ATGGGAAGAGGAAGAGTAGAGCTGAAGAGGATAGAGAACAAAATCAACAGACAAGTAACGTTTGCAAAGCGTAGGAACGGTTTGTGAAG 476  
 C T A A T C A 428  
 M O R O R V E L K R I E N K I N R O V T F A K R R N G L L K 30  
 AAAGCTTATGAATTGTCTGTTCTGTGATGCTGAAGTTGCTCTCATCATCTTCTCCAACCGTGGAAGCTCTATGAGTTTTCAGCTCC 566  
 GC T C CT G C C A 518  
 K A Y E L S V L C D A E V A L I F S N R G K L Y E F C S S 60  
 S V T  
 TCAAACATGCTCAAGACACTTGATCGGTACCAGAAATGCAGCTATGGATCCATTGAAGTCAACAACAAACCTGCCAAGAAGCTTGAGAAC 656  
 C G AA T G T C T G 608  
 S N M L K T L D R Y Q K C S Y G S I E V N N K P A K E L E N 90  
 E  
 AGCTACAGAGAATATCTGAAGCTTAAGGGTAGATATGAGAACCTTCAACGTCAACAGAGAAAATCTTCTTGGGGAGGATTAGGACCTTTG 746  
 G CT G A A T G G A CT C 698  
 S Y R E Y L K L K G R Y E N L Q R Q Q R N L L G E D L G P L 120  
 AATTCAAAGGAGTTAGAGCAGCTTGAGCGTCAACTGGACGGCTCTCTCAAGCAAGTTCGGTCCATCAAGACACAGTACATGCTTGACCAG 836  
 C A G CG T 788  
 N S K E L E Q L E R Q L D G S L K Q V R S I K T Q Y M L D Q 150  
 C  
 CTCTCGGATCTTCAAAAATAAGAGCAAAATGTTGCTTGAACCAATAGAGCTTTGGCAATGAAGCTGGATGATATGATTGGTGTGAGAAGT 926  
 T GG G T C T A TG C T A C C CA 878  
 L S D L Q N K E Q M L L E T N R A L A M K L D D M I G V R S 180  
 G I D A S E H  
 CATCATATG---GGAGGATGGGAAGGCGGTGAA---CAGAATGTTACCTACGCGCATCATCAAGCTCAGTCTCAGGGACTATACCAGCCT 1010  
 C AGGA T TCAA A G T GA C G T AT 968  
 H H M - G G W E G G E - Q N Y T Y A H H Q A Q S Q G L Y Q P 208  
 I G D Q I A G P H S 210  
 CTTGAATGCAATCCAACCTCTGCAAATGGGGTATGATAATCCAGTATGCTCTGAGCAAATCACTGCGACAACACAAGCTCAGGCGCAGCCG 1100  
 TG C T T A AGCC G A GG T GGTG G T C A AA 1058  
 L E C N P T L Q M G Y D N P V C S E Q I T A T T Q A Q A Q P 238  
 D I S H M A V V G S Q 240  
 GGAAACGGTTACATTCCAGGATGGATGCTCTGAGAATCATGTACTGTGATGAAGCTCACCCACAAAAGACCTTATATATATATAAGTAT 1190  
 C C T C G GCGATACTTCTCCCCCAATAAGATCTTAAGCAAGTACTGGTGGGGTCTTCGTGGT 1148  
 G N G Y I P G M M L End 248  
 250  
 (2) AGATACAAGACTTGGATTTGTAGACATAAGTGGCTAATATAATGGTCTCTGAGGATCTTCTAGACATTTGTATCTTTTGGGAATCCTT 1277  
 GCTTATATTAAGAATTC 1294  
 (4) GTGATCTTAGATCTTATGCATATCAATAATAATGTTATTGCACAAGACTTTTGTCTTTGTAGACACAAGTGGCTATAGCTGTAATAG 1235  
 CCTTCAACATCTCTCTTCTGTTTCAGGATTTGTTTGTGCCTATTGTAATTGCTTATATATGTATGTTTGTATAATGTGTGAAATGT 1322  
 S S  
 TAACATCGACCATGTCTCATCTGGTGA<sub>n</sub>  
 S S

Figure 4

30/43

Sequence Range: -12 to 815

38  
CCCGGATCCA AAATGGGAAG AGGGAGAGTA GAATTGAAGA GGATAGAGAA CAAGATCAAT  
K M G R G R V E L K R I E N K I N>

88  
AGGCAAGTGA CGTTTGCAAA GAGAAGGAAT GGTCTTTTGA AGAAAGCATA CGAGCTTTCA  
R Q V T F A K R R N G L L K K A Y E L S>

138  
GTTCTATGTG ATGCGGAAGT TGCTCTCATC ATCTTCTCAA ATAGAGGAAA GCTGTACGAG  
V L C D A E V A L I I F S N R G K L Y E>

188  
TTTTGCAGTA GTTCGAGCAT GCTTCGGACA CTGGAGAGGT ACCAAAAGTG TAACTATGGA  
F C S S S S M L R T L E R Y Q K C N Y G>

238 288  
GCACCAGAAC CCAATGTGCC TTCAAGAGAG GCCTTAGCAG AACTTAGTAG CCAGCAGGAG  
A P E P N V P S R E A L A E L S S Q Q E>

338  
TATCTCAAGC TTAAGGAGCG TTATGACGCC TTACAGAGAA CCCAAAGGAA TCTGTTGGGA  
Y L K L K E R Y D A L Q R T Q R N L L G>

388  
GAAGATCTTG GACCTCTAAG TACAAAGGAG CTTGAGTCAC TTGAGAGACA GCTTGATTCT  
E D L G P L S T K E L E S L E R Q L D S>

438  
TCCTTGAAGC AGATCAGAGC TCTCAGGACA CAGTTTATGC TTGACCAGCT CAACGATCTT  
S L K Q I R A L R T Q F M L D Q L N D L>

488  
CAGAGTAAGG AACGCATGCT GACTGAGACA AATAAACTC TAAGACTAAG GTTAGCTGAT  
Q S K E R M L T E T N K T L R L R L A D>

538 588  
GGGTATCAGA TGCCACTCCA GCTGAACCTT AACCAAGAAG AGGTTGATCA CTACGGTCGT  
G Y Q M P L Q L N P N Q E E V D H Y G R>

638  
CATCATCATC AACAACAACA AACTCCCAA GCTTTCTTCC AGCCTTTGGA ATGTGAACCC  
H H H Q Q Q Q H S Q A F F Q P L E C E P>

688  
ATTCTTCAGA TCGGGTATCA GGGGCAACAA GATGGAATGG GAGCAGGACC AAGTGTGAAT  
I L Q I G Y Q G Q Q D G M G A G P S V N>

738  
AATTACATGT TGGGTTGGTT ACCTTATGAC ACCAACTCTA TTTGAATCTT TCTCACTTAA  
N Y M L G W L P Y D T N S I \* I F L T \*>

788  
TCAATCCCTC TCTTTTTTTT TTTGACATTT TTAAGATGAT GTTTCTA  
S I P L F F F L T F L R \* C F X>

Fig. 5

Sequence Range: -1699 to 3669

-1650  
GAATTCCCCG GATCTCCATA TACATATCAT ACATATATAT AGTATACTAT CTTTAGACTG  
CTTAAGGGGC CTAGAGGTAT ATGTATAGTA TGTATATATA TCATATGATA GAAATCTGAC

-1600  
ATTTCTCTAT ACACTATCTT TTAAGTTATG TATCGTTTCA AAAGTCAGGA CGTACATGTT  
TAAAGAGATA TGTGATAGAA AATTGAATAC ATAGCAAAGT TTTGAGTCCT GCATGTACAA

-1550  
TTAAATTTGG TTATATAACC ACGACCATTT CAAGTATATA TGTCATACCA TACCAGATTT  
AATTTAAACC AATATATTGG TGCTGGTAAA GTTCATATAT ACAGTATGGT ATGGTCTAAA

-1500  
AATATAACTT CTATGAAGAA AATACATAAA GTTGGATTAA AATGCAAGTG ACATCTTTTT  
TTATATTGAA GATACTTCTT TTATGTATTT CAACCTAATT TTACGTTTAC TGTAAGAAAA

-1450  
AGCATAGGTT CATTGTCAT AGAAGAAATA TATAACTAAA AATGAACCTT AACTTAAATA  
TCGTATCCAA GTAAACCGTA TCTTCTTTAT ATATTGATTT TACTTGAAA TTGAATTTAT

-1400  
GATTTTACTA TATTACAATT TTTTCTTTTT ACATGGTCTA ATTTATTTTT CTAAATTAG  
CTAAATGAT ATAATGTAA AAAAGAAAAA TGTACCAGAT TAAATAAAAA GATTTTAAAT

-1350  
TATGATTGTT GTTTTGATGA AACAATAATA CCGTAAGCAA TAGTTGCTAA AAGATGTCCA  
ATACTAACAA CAAAGTACT TTGTATTAT GGCATTCGTT ATCAACGATT TTCTACAGGT

-1300  
AATATTTATA AATTACAAAG TAAATCAAAT AAGGAAGAAG ACACGTGGAA AACACCAAAT  
TTATAAATAT TTAATGTTTC ATTTAGTTTA TTCCTTCTTC TGTGCACCTT TTGTGGTTTA

-1250  
AAGAGAAGAA ATGGAGAAAA CAGAAAGAAA TTTTTTAAAC AGAAAAATCA ATTAGTCTCT  
TTCTCTTCTT TACCTTTTTT GTCTTCTTTT AAAAAATTGT TCTTTTATG TAATCAGGAG

-1200  
AAACCTGAGA TATTTAAAGT AATCAACTAA AACAGGAACA CTTGACTAAC AAAGAAATTT  
TTTGGACTCT ATAAATTCA TTAGTTGATT TTGTCCTTGT GAACTGATTG TTTCTTTAAA

-1150  
GAAATGTGGT CCAACTTICA CTTAATTATA TTGTTTTCTC TAAGGCTTAT GCAATATATG  
CTTTACACCA GGTGAAAGT GAATTAATAT AACAAAAGAG ATTCCGAATA CGTTATATAC

-1100  
CCTTAAGCAA ATGCCGAATC TGTTTTTTTT TTTGTTTATT GGATATTGAC TGAAATAAG  
GGAATTCGTT TACGGCTTAG ACAAAAAAAA AAAACAATAA CCTATAACTG ACTTTTATTC

-1050  
GGGTTTTTTC AACTTGAAG ATCTCAAAG AGAAAACTAT TACAACGGAA ATTCATTGTA  
CCCAAAAAAG TGTGAACCTC TAGAGTTTTC TCTTTTGATA ATGTTGCCTT TAAGTAACAT

-1000  
AAGAAGTGA TTAAGCAAAT TGAGCAAAGG TTTTATGTG GTTTATTTCA TTATATGATT  
TTTCTTCACT AATTCGTTTA ACTCGTTTCC AAAAATACAC CAAATAAAGT AATATACTAA

-950  
GACATCAAAT TGTATATATA TGTTGTTTTT ATTTAACAAT ATATATGGAT ATAACGTACA  
CTGTAGTTTA ACATATATAT ACCAACAAAA TAAATTGTTA TATATACCTA TATTGCATGT

-900  
GACATCAAAT TGTATATATA TGTTGTTTTT ATTTAACAAT ATATATGGAT ATAACGTACA  
CTGTAGTTTA ACATATATAT ACCAACAAAA TAAATTGTTA TATATACCTA TATTGCATGT

-850  
GACATCAAAT TGTATATATA TGTTGTTTTT ATTTAACAAT ATATATGGAT ATAACGTACA  
CTGTAGTTTA ACATATATAT ACCAACAAAA TAAATTGTTA TATATACCTA TATTGCATGT

-800

Fig. 6a

-750

AACTAAATAT GTTTGATTGA CGAAAAAATA TATATGTATG TTTGATTAAAC AACATAGCAC  
TTGATTTATA CAAACTAACT GCTTTTTTTT ATATACATAC AAACATAATTG TTGTATCGTG

-700

ATATTCAACT GATTTTGTGC CTGATCATCT ACAACTTAAT AAGAACACAC AACATTGAAA  
TATAAGTTGA CTAAAAACAG GACTAGTAGA TGTTGAATTA TTCTTGTGTG TTGTAACTTT

-650

AAATCTTTGA CAAAATACTA TTTTGGGTT TGAAATTTTG AATACTTACA ATTATCTTTC  
TTTAGAACT GTTTTATGAT AAAAACCCAA ACTTTAAAC TTATGAATGT TAATAAGAAG

-600

TCGATCTTCC TCTCTTTCCT TAAATCCTGC GTACAAATCC GTCGACGCA TACATTACAC  
AGCTAGAAGG AGAGAAAGGA ATTTAGGACG CATGTTTAGG CAGCTGCGTT ATGTAATGTG

-550

AGTTGTCAAT TGGTTCTCAG CTCTACCAAA AACATCTATT GCCAAAAGAA AGGTCTATTT  
TCAACAGTTA ACCAAGAGTC GAGATGGTTT TTGTAGATAA CGGTTTTCTT TCCAGATAAA

-500

-450

GTACTTCACT GTTACAGCTG AGAACATTAA ATATAATAAG CAAATTGAT AAAACAAAGG  
CATGAAGTGA CAATGTCGAC TCTTGTAAAT TATATTATTC GTTTAAACTA TTTTGTTC

-400

GTTCTCACCT TATTCCAAAA GAATAGTGTA AAATAGGGTA ATAGAGAAAT GTTAATAAAA  
CAAGAGTGA ATAAGGTTTT CTTATCACAT TTTATCCCAT TATCTCTTTA CAATTATTTT

-350

GGAAATAAA AATAGATATT TTGGTTGGTT CAGATTTTGT TTCGTAGATC TACAGGGAAA  
CCTTTAATTT TTATCTATAA AACCAACCAA GTCTAAAAA AAGCATCTAG ATGTCCTTTT

-300

TCTCCGCCGT CAATGCAAAG CGAAGGTGAC ACTTGGGGAA GGACCAGTGG TCCGTACAAT  
AGAGGCGGCA GTTACGTTTC GCTTCCACTG TGAACCCCTT CCTGGTCACC AGGCATGTTA

-250

GTTACTTACC CATTTCTCTT CACGAGACGT CGATAATCAA ATTGTTTATT TTCATATTTT  
CAATGAATGG GTAAAGAGAA GTGCTCTGCA GCTATTAGTT TAACAAATAA AAGTATAAAA

-200

-150

TAAGTCCGCA GTTTTATTAA AAAATCATGG ACCCGACATT AGTACGAGAT ATACCAATGA  
ATTGAGGCGT CAAAATAATT TTTAGTACC TGGGCTGTAA TCATGCTCTA TATGGTTACT

-100

GAAGTCGACA CGCAATCCT AAAGAAACCA CTGTGGTTTT TGCAACAAG AGAAACCAGC  
CTTCAGCTGT GCGTTTAGGA TTTCTTTGGT GACACCAAAA ACGTTTGTTT CTTTGGTTCG

-50

TTTAGCTTTT CCCTAAAACC ACTCTTACCC AAATCTCTCC ATAAATAAAG ATCCCGAGAC  
AAATCGAAAA GGGATTTTGG TGAGAATGGG TTTAGAGAGG TATTTATTTT TAGGGCTCTG

1

TCAAACACAA GTCTTTTAT AAAGGAAAGA AAGAAAACT TTCCTAATTG GTTCATACCA  
AGTTTGTGTT CAGAAAAATA TTTCTTTCTT TTCTTTTGA AAGGATTAAC CAAGTATGGT

51

AAGTCTGAGC TCTTCTTTAT ATCTCTCTTG TAGTTTCTTA TTGGGGGTCT TTGTTTTGTT  
TTCAGACTCG AGAAGAAATA TAGAGAGAAC ATCAAAGAAT AACCCCAAGA AACAAAAACA

101

151

TGTTTCTTTT AGAGTAAGAA GTTCTTAAA AAAGGATCAA AAATGGGAAG GGGTAGGGTT

Fig. 6b

ACCAAGAAAA TCTCATTCTT CAAAGAATTT TTTCTAGTT TTTACCCTTC CCCATCCCAA  
 201  
 CAATTGAAGA GGATAGAGAA CAAGATCAAT AGACAAGTGA CATTCTCGAA AAGAAGAGCT  
 GTTAACTTCT CCTATCTCTT GTTCTAGTTA TCTGTTCACT GTAAGAGCTT TTCTTCTCGA  
 251  
 GGTCTTTTGA AGAAAGCTCA TGAGATCTCT GTTCTCTGTG ATGCTGAAGT TGCTCTTGTT  
 CCAGAAAAC TCTTTCGAGT ACTCTAGAGA CAAGAGACAC TACGACTTCA ACGAGAACAA  
 301  
 GTCTTCTCCC ATAAGGGGAA ACTCTTCGAA TACTCCACTG ATTCTTGATA ACTTCAACTA  
 CAGAAGAGGG TATTCCCCTT TGAGAAGCTT ATGAGGTGAC TAAGAACCAT TGAAGTTGAT  
 351 401  
 ATTCTTTACT TTTAAAAAA TCTTTTAATC TGCTACTTTA TATAGTTTTT TTCCCCCTTA  
 TAAGAAATGA AAATTTTTTT AGAAAATTAG ACGATGAAAT ATATCAAAAA AAGGGGGAAT  
 451  
 AGTTGACTAC TTGATTTGCC CTAATTATTC ACTACTGCTT TTGTTATATA TTTTCTAGGG  
 TCAACTGATG AACTAAACGG GATTAATAAG TGATGACGAA AACAATATAT AAAAGATCCC  
 501  
 CTTCCATTTT TGGATTTTTT GATTAGCCAG AAAAATGTTT AATACAAATT TGTATAATTT  
 GAAGGTAAAA ACCTAAAAAA CTAATCGGTC TTTTACAAA TTATGTTTAA ACATATTAAA  
 551  
 AAAAATCAAA ACTTTAGGGC CGTAGTGAAG TGAACCCTAG AACACACAGA TTATACCATA  
 TTTTtagttt TGAAATCCCG GCATCACTTC ACTTGGGATC TTGTGTGTCT AATATGGTAT  
 601  
 GTAATTACCT TGATATATTG TGCAATATTT ATCAGCATCA TATCTTCAAA CTCAAGAGAT  
 CATTAAATGGA ACTATATAAC ACGTTATAAA TAGTCGTAGT ATAGAAGTTT GAGTTCTCTA  
 651 701  
 ATAGAAGGGT ATGTTAATCT TTGAAGTAGG GTTTTGATCC CTAACTCATA ATGAATCCTT  
 TATCTTCCCA TACAATTAGA AACTTGATCC CAAAAGTAGG GATTGAGTAT TACTTAGGAA  
 751  
 TTGTTCTCCA ATAGCCATGT CTTTCGAATT TGCAGATCTA AGCTCTAATT GATGCCATAG  
 AACAAAGAGG TATCGGTACA GAAAGCTTAA ACGTCTAGAT TCGAGATTAA CTACGGTATC  
 801  
 TAAGAAAATA AGATCTGTAG TTTTCACTCG CTCAGTGAAT TCGAGTTTAA AATGAAGTGT  
 ATTCTTTTAT TCTAGACATC AAAAGTGAGC GAGTGACTCA AGCTCAAAAT TTACTTCACA  
 851  
 CGTTTCTTTT TTCATATATA GTTGCAACTG GATTATAATT AAAAAATATT ATGGGACGAG  
 GCAAAGAAAA AAGTATATAT CAACGTTGAC CTAATATTAA TTTTATATA TACCGTGCTC  
 901  
 AAAATAATTT AAAATAGATA TAGATAACAA TGTCAAATTG AGAATTTTTT ATTAGAAAGA  
 TTTTATTAAA TTTTATCTAT ATCTATTGTT ACAGTTTAA TCTTAAAAAA TAATCTTTCT  
 951 1001  
 ATATTTAACT TACGAGTTGT TTTTTCAG CTGTAAAAGA ATATCTAATT TGTTCTCAGC  
 TATAAATTGA ATGCTCAACA AAAAAAGTC GACATTTTCT TATAGATTAA ACAAGAGTGC  
 1051  
 ACTGTGTCTT CATGTTTTGC AAATCTAAGC AAAGAAAATG TTTAAACTCG GATCTTAAGA  
 TGACACAGAA GTACAAAACG TTAGATTTCG TTTCTTTTAC AAATTTGAGC CTAGAATTCT

Fig. 6c

09869582.022802

1101  
TTATGAACTC GTAATATAAA ACACTATATA GTATTAAATT TGAAC TAGTG TTGCTTCTTT  
AATACTTGAG CATTATATTT TGTGATATAT CATAATTTAA ACTTGATCAC AACGAAGAAA

1151  
TGCTACTTTG ACTTTAGAAA TTAAAACTGA AACAAAGATG TCAAATCTGA GTAGGGAGTC  
ACGATGAAAC TGAAATCTTT AATTTTGACT TTGTTTCTAC AGTTTAGACT CATCCCTCAG

1201  
TTTGACCTCT GGGGATCCAT AAAAAGAACT AACTCCATCC TAAATCGGC TTCTTACCGA  
AAACTGGAGA CCCCTAGGTA TTTTCTTGA TTGAGGTAGG ATTTTAGCCG AAGAATGGCT

1251 1301  
TGGTCAAAC TAGCTCCAAC AAGCAACAGC TGTTCCTCTT TTTTCTTTT TTTTCTTTT  
ACCAGTTTGA ATCGAGGTTG TTCGTTGTCG ACAAGAAGAA AAAAAAAAAA AAAAAAAAAA

1351  
TTTAAGCATT GTCCTTGTTT TGAAAAAAA TAAGATTGGT AAATTGGCAA GATTATAATA  
AAATTCGTAA CAGGAACAAG ACTTTTTTTT ATTCTAACCA TTTAACCGTT CTAATATTAT

1401  
ATTTATTATA ATGTGTCGCA CTAAGAAGAT TTTCTGTACC TAATTGTAGC AAAATTAAAG  
TAAATAATAT TACACAGCGT GATTCTTCTA AAAGACATGG ATTAACATCG TTTTAATTTT

1451  
AAACCGCAGT TAGAACTCGA AGCTAAGAGC ATAGGGTCTA TGATTCATAC TGTTTTGTTA  
TTTGGCGTCA ATCTTGAGCT TCGATTCTCG TATCCCAGAT ACTAAGTATG ACAAACAAT

1501  
TTATAAAGGT ATCATAGAGA TCGGTACTTG ATTTGTTATA GGAAATCTTG GTTTAATTGC  
AATATTCCA TAGTATCTCT AGCCATGAAC TAAACAATAT CCTTTAGAAC CAAATTAACG

1551 1601  
ATAAAACCAT CATTAGATTT ATCCTAAAAT GTGATGATAT TTTGGTCACA TCTCCATATT  
TATTTTGGTA GTAATCTAAA TAGGATTTTA CACTACTATA AAACCAAGTG AGAGGTATAA

1651  
ATTTATATAA TAAATGATA ATTGGTTGAT GATAAAGCTA ACCCTAATTC TGTGAAATGA  
TAAATATATT ATTTTACTAT TAACCAACTA CTATTTGAT TGGGATTAAG ACACCTTACT

1701  
TCAGTATGGA GAAGATACTT GAACGCTATG AGAGGTACTC TTACGCCGAA AGACAGCTTA  
AGTCATACCT CTCTATGAA CTTGCGATAC TCTCCATGAG AATGCGGCTT TCTGTGGAAT

1751  
TTGCACCTGA GTCCGACGTC AATGTATTTT AATAAATATT TCTCCTTTTA ATCCACATAT  
AACGTGGACT CAGGCTGCAG TTACATAAAG TTATTTATAA AGAGGAAAAT TAGGTGTATA

1801  
ATATTATATC AATCTATTTG TAGTATTGAT GAATTTTATT TGTATAAAAC TTCTGGTACA  
TATAATATAG TTAGATAAAC ATCATAACTA CTTAAAATAA ACATATTTTG AAGACCATGT

1851 1901  
CAGACAAACT GGTGATGGA GTATAACAGG CTTAAGGCTA AGATTGAGCT TTTGGAGAGA  
GTCTGTTTGA CCAGCTACCT CATATTGTCC GAATTCGGAT TCTAACTCGA AAACCTCTCT

1951  
AACCAGAGGT ACACATTTAC ACTCATCACA TTTCTATCTA GAAATCGAT CGGGTTCCAT  
TTGGTCTCCA TGTGTAAATG TGAGTAGTGT AAAGATAGAT CTTTGTAGCT GCCCAAGGTA

2001  
TTTAAAGTAA GTTAAATTC ATTGATGCTA TTGAAATTCA GGCATTATCT TGGGAAGAC

Fig. 6d

AAATTTTCATT CAATTTTAAAG TAACTACGAT AACTTTAAGT CCGTAATAGA ACCCCTTCTG

2051  
TTGCAAGCAA TGAGCCCTAA AGAGCTTCAG AATCTGGAGC AGCAGCTTGA CACTGCTCTT  
AACGTTTCGTT ACTCGGGATT TCTCGAAGTC TTAGACCTCG TCGTCGAACT GTGACGAGAA

2101  
AAGCACATCC GCACTAGAAA AGTATTGCCT TCTGCTATTT CGTTGAACAT ATCTATATAA  
TTCGTGTAGG CGTGATCTTT TCATAACGGA AGACGATAAA GCAACTTGTA TAGATATATT

2151 2201  
CTTAAACGTT TACAAGTGTT ATTATAATGT GAACATTGAA ATACATATGT GTATGTATCA  
GAATTTGCAA ATGTTCAACAA TAATATTACA CTTGTAACTT TATGTATACA CATACATAGT

2251  
ATATATATAT CAGTAATCAA TATCAATTTG ATATGTCTAT AGGTTGGTTC GAATGTATGA  
TATATATATA GTCATTAGTT ATAGTTAAAC TATACAGATA TCCAACCAAG CTTACATACT

2301  
GTTATGTTGT GTATTTTAAAG ACTCCATATT ACTTAAAGTA ATGGGTTGTT AATGTTGATG  
CAATACAACA CATAAAATTC TGAGGTATAA TGAATTTTCAT TACCAACAA TTACAACCTAC

2351  
TGTGTGTATG CAGAACCAAC TTATGTACGA GTCCATCAAT GAGCTCCAAA AAAAGGTATG  
ACACACATAC GTCTTGTTG AATACATGCT CAGGTAGTTA CTCGAGGTTT TTTTCCATAC

2401  
TAAACCCCT ATCAAATGTA TGTCTTATAG AGAAACGTAT AGGAAAGCTA ATTAACAATC  
ATTTTGGGGA TAGTTTACAT ACAGAATATC TCTTTGCATA TCCTTTTCGAT TAATTGTTAG

2451 2501  
GTGCCGTTTC GGAAATGACA GGAGAAGGCC ATACAGGAGC AAAACAGCAT GCTTTCTAAA  
CACGGCAAAG CCTTTACTGT CCTCTTCCGG TATGTCCTCG TTTTGTCTGA CGAAAGATT

2551  
CAGGTAACAC ATGTCATCAT TTCTCTTTCA TCAACATGTT GTCCATTGCA TTACTGTTAC  
GTCCATTGTG TACAGTAGTA AAGAGAAAGT AGTTGTACAA CAGGTAACGT AATGACAATG

2601  
CTTCCACTGT TCTGCTCCAC ACTTCCAGCC AAGCTATACC TACGATATCT TCATATCTCC  
GAAGGTGACA AGACGAGGTG TGAAGGTCGG TTCGATATGG ATGCTATAGA AGTATAGAGG

2651  
ACTTAACTTC GGCACCATTA AATAAAAATA GAAAATCTTT GCAAATTTGT TTGAAATAGC  
TGAATTGAAG CCGTGGTAAT TTATTTTTAT CTTTTAGAAA CGTTTAAACA AACTTTATCG

2701  
ATAGATGTTG TCTATTGATT GATATAATCA CCAGCCTGTA CGTAGATATG GTTTGTCCGT  
TATCTACAAC AGATAACTAA CTATATTAGT GGTCCGACAT GCATCTATAC CAAACAGGCA

2751 2801  
TTAGTTTTAA GGTGTCTCTC GGATTGAAAA TATTTTGAAA TCTTTTGAAA TGTGTGCCCC  
AATCAAAATT CCACAGAGAG CCTAACTTTT ATAAACTTT AGAAACTTT ACAAACAGGG

2851  
ATCATCTTA CTTAGCTCAT ATCTATGTAT ATGAATATAG ACACTACTCC TAATTATAAA  
TAGTAAGAAT GAATCGAGTA TAGATACATA TACTTATATC TGTGATGAGG ATTAATATTT

2901  
ATGTTATAAT AGTTCATTGC ATGAGTGCAA CTGTGAAAAA AACTATTTGT AACCATTGCA  
TACAATATTA TCAAGTAACG TACTCACGTT GACACTTTTA TTGATAAACA TTGGTAACGT

Fig. 6e

09869582-022302

2951  
TATATATAGT TTCTTCACTT TGAAAATTGA TGATGATAAT ATGGTTTGAA ATAAATTTGC  
ATATATATCA AAGAAGTGAA ACTTTTAACT ACTACTATTA TACCAACTT TATTTAAACG

3001  
TGCGAGATCA AGGAGAGGGA AAAAATTCTT AGGGCTCAAC AGGAGCAGTG GGATCAGCAG  
ACCGTCTAGT TCCTCTCCCT TTTTAAAGAA TCCCAGATTG TCCTCGTCAC CCTAGTCGTC

3051 3101  
AACCAAGGCC ACAATATGCC TCCCCCTCTG CCACCGCAGC AGCACCAAAT CCAGCATCCT  
TTGGTTCCGG TGTTATACGG AGGGGGAGAC GGTGGCGTCG TCGTGGTTTA GGTCTAGGA

3151  
TACATGCTCT CTCATCAGCC ATCTCCTTTT CTCAACATGG GGTAACAAAA AATTACTAAT  
ATGTACGAGA GAGTAGTCGG TAGAGGAAAA GAGTTGTACC CCATTGTTTT TTAATGATTA

3201  
CAGTCTTAAT TTAAAGCACA TATGTTATGC AAGCTAGTTA CGTTAGGTGT TGTAAATTTCA  
GTCAGAATTA AATTTCGTGT ATACAATACG TTCGATCAAT GCAATCCACA ACATTAAAGT

3251  
TTGAAGTTAT AGCTGTTAGT GATGGTTACA TGATGCTAGA TTTTGAACT AGAAAACTTT  
AACTTCAATA TCGACAATCA CTACCAATGT ACTACGATCT AAAACTTTGA TCTTTTGAA

3301  
ATTTTAAAC ATTATTTTAT TAACGTAGGT TAATGCAATG GTCGCCAAAC GAACAACTT  
TAAATTTTG TAATAAAATA ATTGCATCCA ATTACGTTAC CAGCGGTTTG CTTGTTTGAA

3351 3401  
ATTAGTGTGG AAAAATGTAC ATGGAATGGT TCGGAAAAGC CTAAGTCGAC TTTTGTGTGTT  
TAATCACACC TTTTACATG TACCTTACCA ACGCTTTTCG GATTACAGTG AAAACAACAA

3451  
GTTGGTCTAT GTGTTTAAAGT ACAATTTTAG TTTGTTAGAT AAATGAAATT AATATATCTT  
CAACCAGATA CACAAATCA TGTTAAATC AAACAATCTA TTTACTTTAA TTATATAGAA

3501  
TGACATTTCA CAATGGACTG ATATTTGATT TTCCTTTGTT GTACGGTGAA ACATATGATT  
ACTGTAAAGT GTTACCTGAC TATAAACTAA AAGGAAACAA CATGCCACTT TGTATACTAA

3551  
ACATATGCAC TTTCATATAT ATCCTATGTA TGATTGTGAA TGCAGTGGTC TGTATCAAGA  
TGTATACGTG AAAGTATATA TAGGATACAT ACTAACACTT ACGTCACCAG ACATAGTTCT

3601  
AGATGATCCA ATGGCAATGA GGAGGAATGA TCTCGAACTG ACTCTTGAAC CCGTTTACAA  
TCTACTAGGT TACCGTTACT CCTCCTTACT AGAGCTTGAC TGAGAACTTG GGCAAATGTT

3651  
CTGCAACCTT GGCTGCTTCG CCGCATGA  
GACGTTGGAA CCGACGAAGC GGCGTACT

Fig. 6f



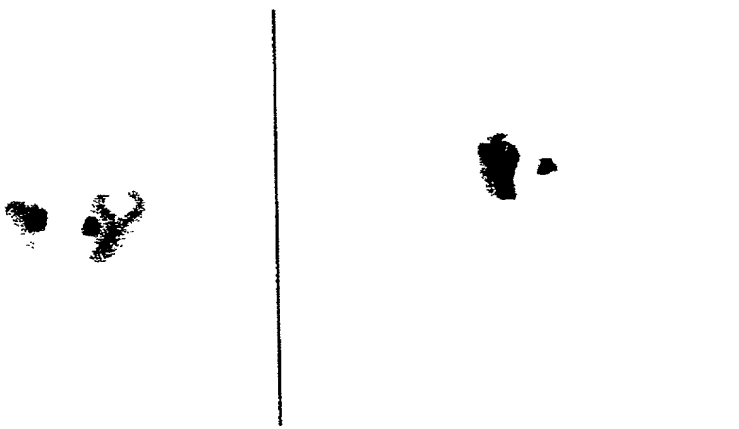


Fig. 9

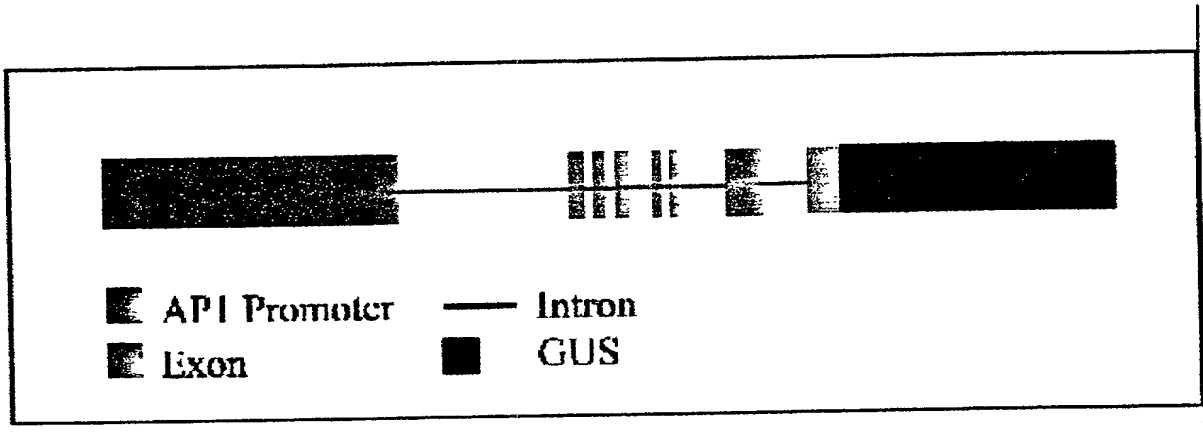


Fig. 7

09869582-022809

Sequence Range: -140 to 1080

```

-91
GAATTCGGCA CGAGAACTTT CCTAATTGGT TCATACCAAA GTCTGAGCTC TTCTTTATAT

-41
CTCTCTTGTA GTTCTTTATT GGGGGTCTTT GTTTTGTITG GTTCTTTTAG AGTAAGAAGT

10
TTCTTAAAAA AGGATCAAAA ATGGGAAGGG GTAGGGTTCA ATTGAAGAGG ATAGAGAACA
M G R G R V Q L K R I E N>

60
AGATCAATAG ACAAGTGACA TTCTCGAAAA GAAGAGCTGG TCTTTTGAAG AAAGCTCATG
K I N R Q V T F S K R R A G L L K K A H>

110
AGATCTCTGT TCTCTGTGAT GCTGAAGTTG CTCTTGTGTG CTCTCTCCAT AAGGGGAAAC
E I S V L C D A E V A L V V F S H K G K>

210
TCTTCGAATA CTCCACTGAT TCTTGTATGG AGAAGATACT TGAACGCTAT GAGAGGTACT
L F E Y S T D S C M E K I L E R Y E R Y>

260
CTTACGCCGA AAGACAGCTT ATTGCACCTG AGTCCGACGT CAATACAAAC TGGTCGATGG
S Y A E R Q L I A P E S D V N T N W S M>

310
AGTATAACAG GCTTAAGGCT AAGATTGAGC TTTTGGAGAG AAACCAGAGG CATTATCTTG
E Y N R L K A K I E L L E R N Q R H Y L>

360
GGGAAGACTT GCAAGCAATG AGCCCTAAG AGCTTCAGAA TCTGGAGCAG CAGCTTGACA
G E D L Q A M S P K E L Q N L E Q Q L D>

410
CTGCTCTTAA GCACATCCGC ACTAGAAAAA ACCAACTTAT GTACGAGTCC ATCAATGAGC
T A L K H I R T R K N Q L M Y E S I N E>

510
TCCAAAAAAA GGAGAAGGCC ATACAGGAGC AAAACAGCAT GCTTTCTAAA CAGATCAAGG
L Q K K E K A I Q E Q N S M L S K Q I K>

560
AGAGGGAAAA AATTCTTAGG GCTCAACAGG AGCAGTGGGA TCAGCAGAAC CAAGGCCAEE
E R E K I L R A Q Q E Q W D Q Q N Q G H>

610
ATATGCCTCC CCCTCTGCCA CCGCAGCAGC ACCAAATCCA GCATCCTTAC ATGCTCTCTC
N M P P P L P P Q Q H Q I Q H P Y M L S>

660
ATCAGCCATC TCCTTTTCTC AACATGGGTG GTCTGTATCA AGAAGATGAT CCAATGGCAA
H Q P S P F L N M G G L Y Q E D D P M A>

710
TGAGGAGGAA TGATCTCGAA CTGACTCTTG AACCCGTTTA CAACTGCAAC CTTGGCTGCT
M R R N D L E L T L E P V Y N C N L G C>

810
TCGCCGCATG AAGCATTTCC ATATATATAT ATTTGTAATC GTCAACAATA AAAACAGTTT
F A A *

860
GCCACATACA TATAAATAGT GGCTAGGCTC TTTTCATCCA ATTAATATAT TTTGGCAAAT

910
GTTCGATGTT CTTATATCAT CATATATAAA TTAGCAGGCT CTTTCTTCT TTTGTAATTT

```

Fig. 8a

960  
GATAAGTTTA TTGCTTCAA TATGGAGCAA AATTGTAATA TATTGAAGG TCAGAGAGAA  
1010  
TGAACGTGAA CTTAATAGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAACCC  
1060  
CGACGTAGCT CGAGGAATTC

Fig. 8b

Sequence Range: -346 to 1028

GAATTCGGGA TTCACAAAAA CTTTCTCTCA GATTCACAAT CTCATCACAA CCCTTCAAAA -297

AGAGAAAAGA TCTAAAGAAT AAACAAGAGC CCTAATATCA AATCACAACC AAAAAACCA -247

AAGAAAGCTA ATTAAAGTTT TCTCTCTAGC TATTCCTCTT CTTTCTTGT TCTTGAAAC -197

TAGGGTTTAC TTCACCAAAA GATAAGATCT TTCCCCAGAA AAAGCAATAC CCAAGTCATG -147

TTTCTGTGTG TCTGTATATA GATAAAACAT TACATACCCT AATAAGGTTA CACAAATAGC -97 -47

TATAAAAGAG GGAAAATAAG ATAGGGATTT TTTGGGGTGA GGAAAGATGG GAAGAGGAAG 4  
M G R G R>

AGTAGAGCTC AAGAGGATAG AGAACAAAAT CAACAGACAA GTGACGTTTG CTAAACGTAG 54  
V E L K R I E N K I N R Q V T F A K R R>

AAATGGTTTG CTGAAAAAAG CTTATGAGCT TTCTGTTCTC TGCGATGCTG AAGTCTCTCT 104  
N G L L K K A Y E L S V L C D A E V S L>

CATCGTCTTC TCCAACCGTG GCAAGCTCTA CGAGTTCTGC AGCACCTCCA ACATGCTCAA 154  
I V F S N R G K L Y E F C S T S N M L K>

GACACTGGAA AGGTATCAGA AGTGTAGCTA TGGCTCCATT GAAGTCAACA ACAAACTGCT 204 254  
T L E R Y Q K C S Y G S I E V N N K P A>

TAAAGAGCTT GAGAACAGCT ACAGAGAGTA CTTGAAGCTG AAAGGTAGAT ATGAAAATCT 304  
K E L E N S Y R E Y L K L K G R Y E N L>

GCAACGTCAG CAGAGAAATC TTCTTGGAGA GGATCTTGGA CCTCTGAATT CAAAGGAGCT 354  
Q R Q Q R N L L G E D L G P L N S K E L>

AGAGCAGCTT GAGCGTCAAC TAGACGGCTC TCTGAAGCAA GTTCGCTGCA TCAAGACACA 404  
E Q L E R Q L D G S L K Q V R C I K T Q>

GTATATGCTT GACCAGCTCT CTGATCTTCA AGGTAAGGAG CATATCTTGC TTGATGCCAA 454  
Y M L D Q L S D L Q G K E H I L L D A N>

CAGAGCTTTG TCAATGAAGC TGGAAGATAT GATCGGCGTG AGACATCACC ATATAGGAGG 504 554  
R A L S M K L E D M I G V R H H H I G G>

AGGATGGGAA GGTGGTGATC AACAGAATAT TGCCTATGGA CATCCTCAGG CTCATTCTCA 604  
G W E G G D Q Q N I A Y G H P Q A H S Q>

GGGACTATAC CAATCTCTTG AATGTGATCC CACTTTGCAA ATTGGATATA GCCATCCAGT 654  
G L Y Q S L E C D P T L Q I G Y S H P V>

GTGCTCAGAG CAAATGGCTG TGACGGTGCA AGGTCAGTCC CAACAAGGAA ACGGCTACAT 704  
C S E Q M A V T V Q G Q S Q Q G N G Y I>

Fig. 10a

09869582-022807

754  
CCCTGGCTGG ATGCTGTGAG CGATACTTCT TCCCCAATA AAGATCTTAA GCAAGTACTG  
P G W M L \*  
804 854  
GTGGGGTCTT CGTGGTGTGA TCTTAGATCT TATGCATATG AATAATAATG TTATTGCACA  
904  
AGACTTTTGC TTTTGTAGAC ACAAGTGGCT ATAGCTGTAA TAGCCTTCAA CATCTCTCTT  
954  
CTGTTTCAGG ATTTGTGTTGT GCCTATTGTA ATTGCTTATA TATGTATGGT TTGTATAATG  
1004  
TGTGAAATGT TAACATCGAC CATGTCTCAT CTGGTGAAAA AAAAAAAAAA AAAA

Fig. 10b

09869582 02200

Sequence Range: -395 to 908

GAATTCGGGC CCTCACACAT TTCTTATCTT TTGCTCTCAA TAGATTCCAT TGATTCAAAA -346  
CAAAATTTTC ATTAAGATTT CACAACCTCC ACACACTTCC AAACACAATT AAAGAGAGGA -296  
AAAAGAATCA ATAACCCTAT AAATAAAAA TCAGACAAAC AGAAGTTTCC TCTTCTTCTT -246  
CCTTAAGCTA GTACCTTTTG TTCTTGAAAT TAGGGTTAAT TTCTTTTTTC CAAATACCAT -196  
CAATTCTCCA GACCATAAAA ACTCAAAAAG ATCAGATCTT TCCTCTGAAA AAGAGATACC -146 -96  
CAACTTATGT TTTTGTGTGT CTGTATATAG ATAAACATTA CATACCCATA TTTGTGTATA -46  
GACATAAAAA GTGGAAATTA AGGTAACAAA AAGAAATGGG AAGAGGAAGA GTAGAGCTGA 5  
M G R G R V E L>  
AGAGGATAGA GAACAAAATC AACAGACAAG TAACGTTTGC AAAGCGTAGG AACGTTTGT 55  
K R I E N K I N R Q V T F A K R R N G L>  
TGAAGAAAGC TTATGAATTG TCTGTCTCT GTGATGCTGA AGTTGCTCTC ATCATCTTCT 105  
L K K A Y E L S V L C D A E V A L I I F>  
CCAACCGTGG AAAGCTCTAT GAGTTTGTGA GCTCCTCAA CATGCTCAAG AACTTTGATC 155 205  
S N R G K L Y E F C S S S N M L K T L D>  
GGTACCAGAA ATGCAGCTAT GGATCCATTG AAGTCAACAA CAAACCTGCC AAAGAAGTTG 255  
R Y Q K C S Y G S I E V N N K P A K E L>  
AGAACAGCTA CAGAGAATAT CTGAAGCTTA AGGGTAGATA TGAGAACCTT CAACGTCAAC 305  
E N S Y R E Y L K L K G R Y E N L Q R Q>  
AGAGAAATCT TCTTGGGGAG GATTTAGGAC CTTTGAATTC AAAGGAGTTA GAGCAGCTTG 355  
Q R N L L G E D L G P L N S K E L E Q L>  
AGCGTCAACT GGACGGCTCT CTCAAGCAAG TTCGGTCCAT CAAGACACAG TACATGCTTG 405  
E R Q L D G S L K Q V R S I K T Q Y M L>  
ACCAGCTCTC GGATCTTCAA AATAAGAGC AAATGTTGCT TGAAACCAAT AGAGCTTTGG 455 505  
D Q L S D L Q N K E Q M L L E T N R A L>  
CAATGAAGCT GGATGATATG ATTGGTGTGA GAAGTCATCA TATGGGAGGA TGGGAAGGCG 555  
A M K L D D M I G V R S H H M G G W E G>  
GTGAACAGAA TGTTACCTAC GCGCATCATC AAGCTCAGTC TCAGGGACTA TACCAGCCTC 605  
G E Q N V T Y A H H Q A Q S Q G L Y Q P>  
TTGAATGCAA TCCAATCTG CAAATGGGGT ATGATAATCC AGTATGCTCT GAGCAAATCA 655  
L E C N P T L Q M G Y D N P V C S E Q I>

Fig. 11a

705  
CTGCGACAAC ACAAGCTCAG GCGCAGCCGG GAAACGGTTA CATTCAGGA TGGATGCTCT  
T A T T Q A Q A Q P G N G Y I P G W M L>  
755 805  
GAGAATCATG TACTGTGATG AAGCTCACCC ACAAAGACC TTATATATAT ATAAAGTATA  
\*  
GATACAAGAC TTGGATTTGT AGACATAAGT GGCTAATATA ATGGTCCTGA GGATCTTCTA  
905  
GACATTTGTA TCTTTTGGGA ATCCTTGCTT ATATTAAGAA TTC

Fig. 11b

09/869582-02200

**DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION  
USING AN APPLICATION DATA SHEET (37 CFR 1.76)**

As the below named inventor(s), I/we declare that:

This declaration is directed to:

☐ The attached application, or

☒ U.S. Patent Application No. 09/869,582, claiming benefit of priority  
under 35 USC § 371 of International Application No. PCT/US99/24407 with International  
Filing Date of October 15, 1999.

☐ as amended on \_\_\_\_\_ (if applicable);

I/we believe that I/we am/are the original and first inventor(s) of the subject matter which is claimed and  
for which a patent is sought;

I/we have reviewed and understand the contents of the above-identified application, including the  
claims, as amended by any amendment specifically referred to above;

I/we acknowledge the duty to disclose to the United States Patent and Trademark Office all information  
known to me/us to be material to patentability as defined in 37 CFR 1.56, including material information  
which became available between the filing date of the prior application and the National or PCT  
International filing date of the continuation-in-part application, if applicable; and

All statements made herein of my/our own knowledge are true, all statements made herein on  
information and belief are believed to be true, and further that these statements were made with the  
knowledge that willful false statements and the like are punishable by fine or imprisonment, or both,  
under 18 U.S.C. 1001, and may jeopardize the validity of the application or any patent issuing thereon.

**FULL NAME OF INVENTOR(S)**

Inventor 1 Martin F. Yanofsky Date: 1/29/02  
Signature: Martin F. Yanofsky Citizen of: United States

Inventor 2 \_\_\_\_\_ Date: \_\_\_\_\_  
Signature: \_\_\_\_\_ Citizen of: \_\_\_\_\_

Inventor 3 \_\_\_\_\_ Date: \_\_\_\_\_  
Signature: \_\_\_\_\_ Citizen of: \_\_\_\_\_

Inventor 4 \_\_\_\_\_ Date: \_\_\_\_\_  
Signature: \_\_\_\_\_ Citizen of: \_\_\_\_\_

☐ Additional inventors are being named on form(s) attached hereto.